In/Outdoor Visual Detection on Embedded

Xinmiao Xiong, Yi Wen Chen, Ziqi Liao, Pin Chun Lu

Outlines

Background:

- Machine Learning at Edge Device

Remind your personal item through smart glasses:

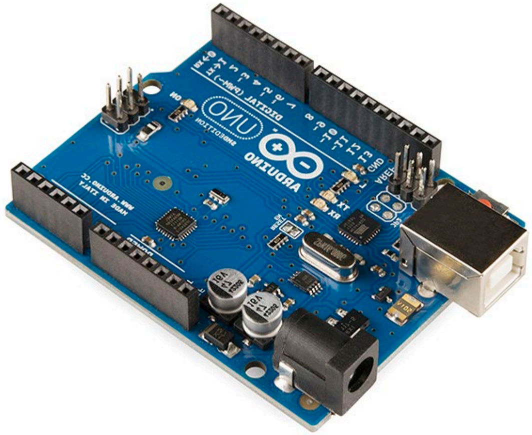
- Indoor Object Detection and Tracking

Wildfire Predicting at Remote Sensing:

- Early Smoke Detection for High-Resolution Images on Edge

Machine Learning at Edge Device

- **Bandwidth**—ML algorithms on edge devices can extract meaningful information from data that would otherwise be inaccessible due to bandwidth constraints.
- **Latency**—On-device ML models can respond in real-time to inputs, enabling applications such as autonomous vehicles, which would not be viable if dependent on network latency.
- **Economics**—By processing data on-device, embedded ML systems avoid the costs of transmitting data over a network and processing it in the cloud.



Object detection in 2D images and videos

What is YOLO? (You Only Look Once)

- Real-time object detection system using deep learning
- Single neural network for object detection and recognition
- Designed for speed and accuracy on images/videos

Original Dataset

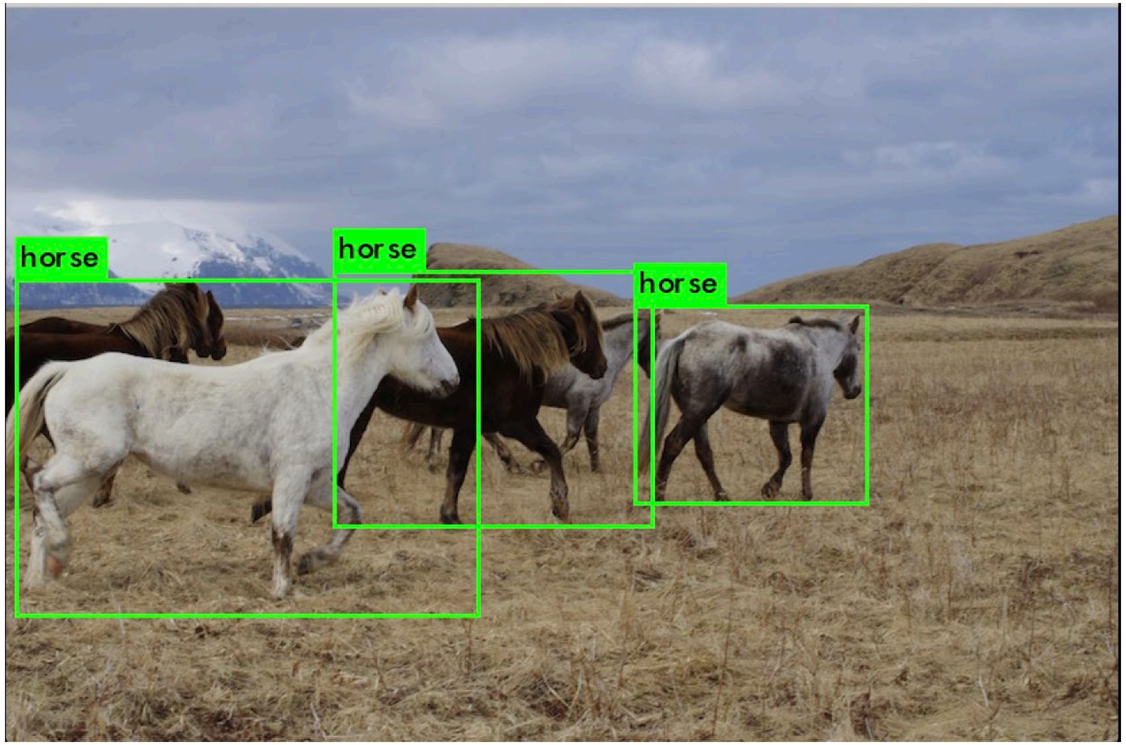
- Trained on COCO (Common Objects in Context) dataset

Object Detection Workflow

- Divides image into grid cells
- Predicts bounding boxes and class probabilities per cell
- Detects multiple objects simultaneously

Pre-trained Weights and Transfer Learning

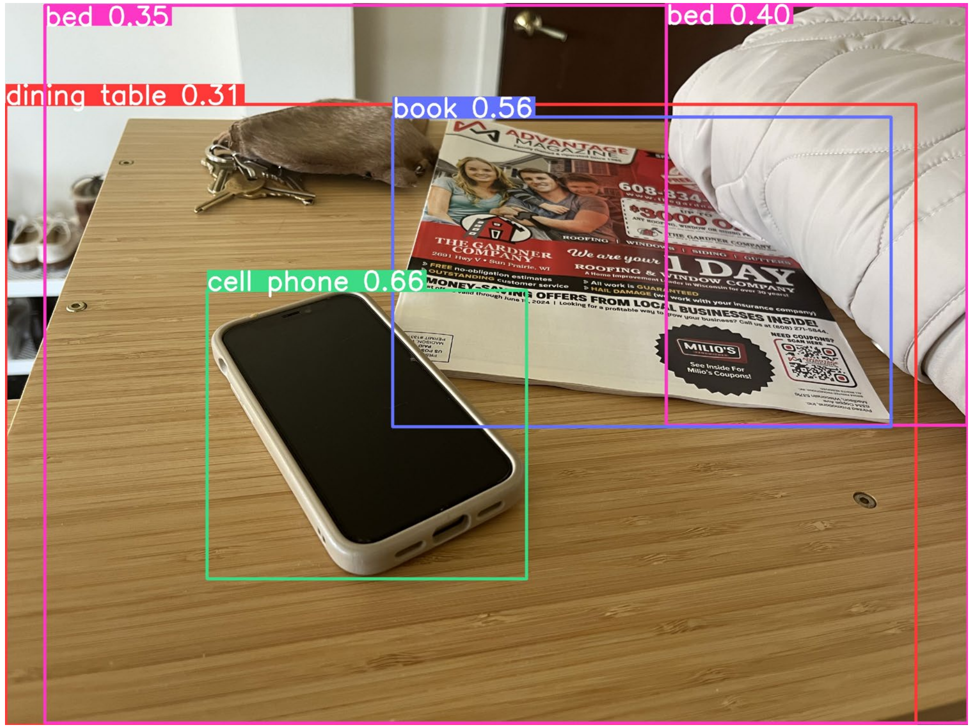
- Pre-trained on large datasets like COCO
- Enables transfer learning for custom object detection



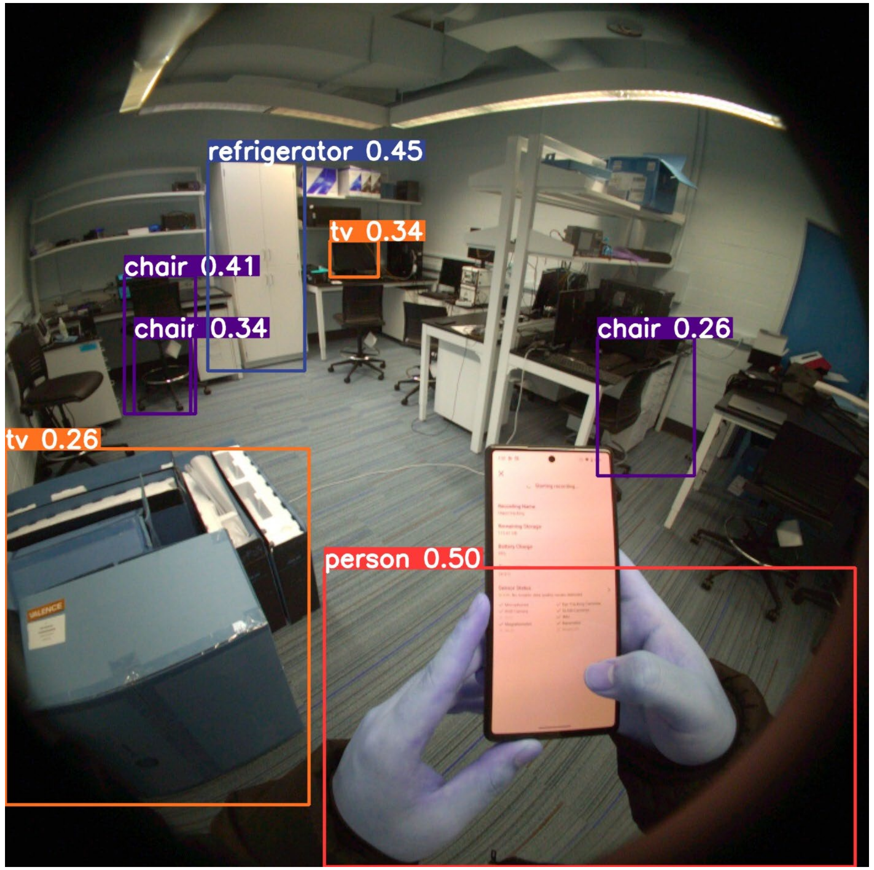
Resource:https://pjreddie.com/darknet/yolo/

Finding Your Item Indoor

A pretrained YOLOv8 model with singe frame image



Find your phone on the table



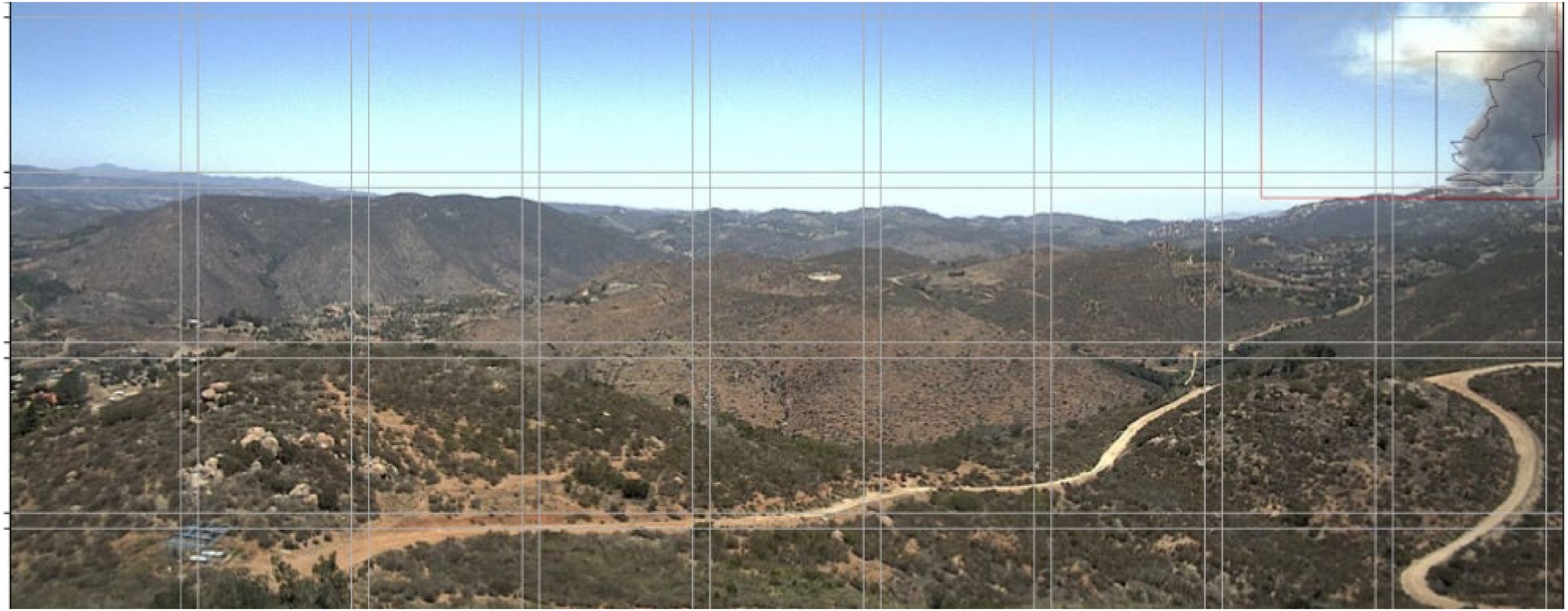
Tracking objects in the room

Overview- Wildland Fire Image Dataset

Sequences of wildland fire images as seen from fixed cameras at HPWREN sites.

Resolution and distance can be varied in a wild range.

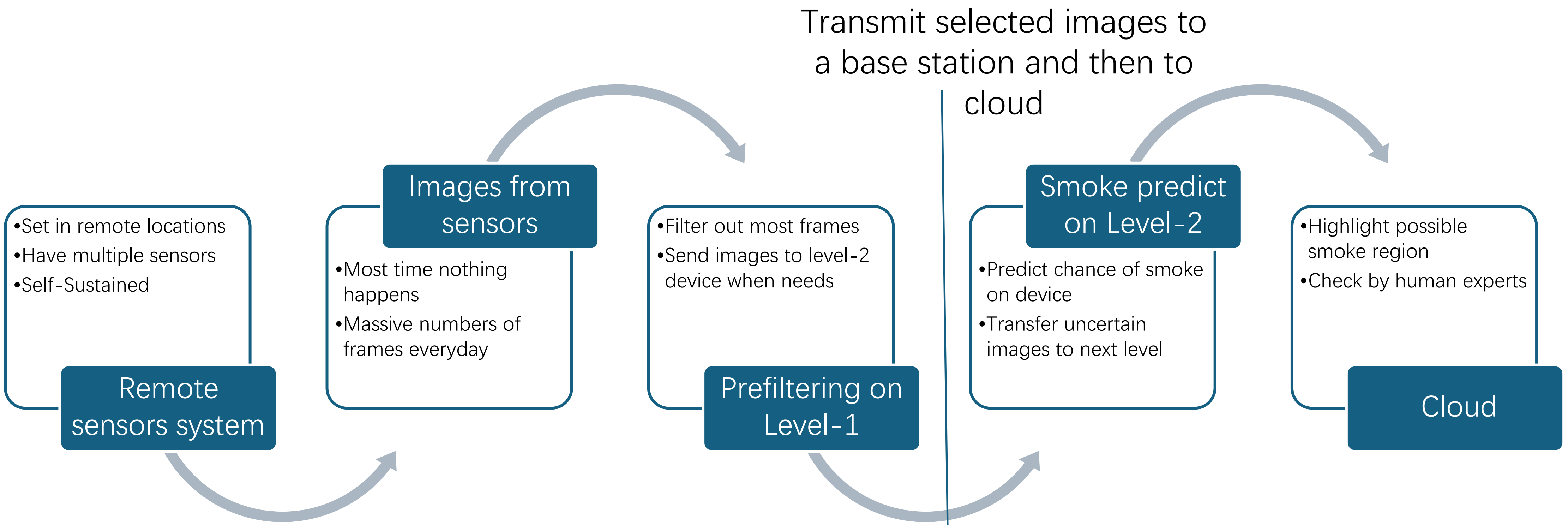
HPWREN Fire Ignition images Library for neural network training (ucsd.edu)



Motivation

- A comprehensive sensing framework is needed to assess wildfire risks more accurately and enable strategic decisions
  - Commonly used sensors enable (temperature, humidity, particle, etc.) **short-range detection.**
  - **Long-range detection** (~1km-10km) can be enabled using **cameras**.
- Challenges in edge-devices:
  - Resource-constraint devices.
  - Communication is costly.
  - For cameras – unexpected conditions (e.g., fog, rain)

Hierarchical Smoke Prediction Framework



These operations will happen on the sensor suit (locally). This is work in progress

The rest will happen at a base station and/or cloud
(Our focus today)

Visual sensors help wildfire detection

Detecting result based on YOLOv8(finetuned)

Benefit of detection on smoke:

1. Location known.
2. Low false positive rate.

Weakness:

1. Hard to track deformable smoke
2. Computing heavy
3. High false negative rate
4. Weak at far-distance smoke

Hard cases in distanced smoke

Process High-resolution at resource-limited device

Input images:

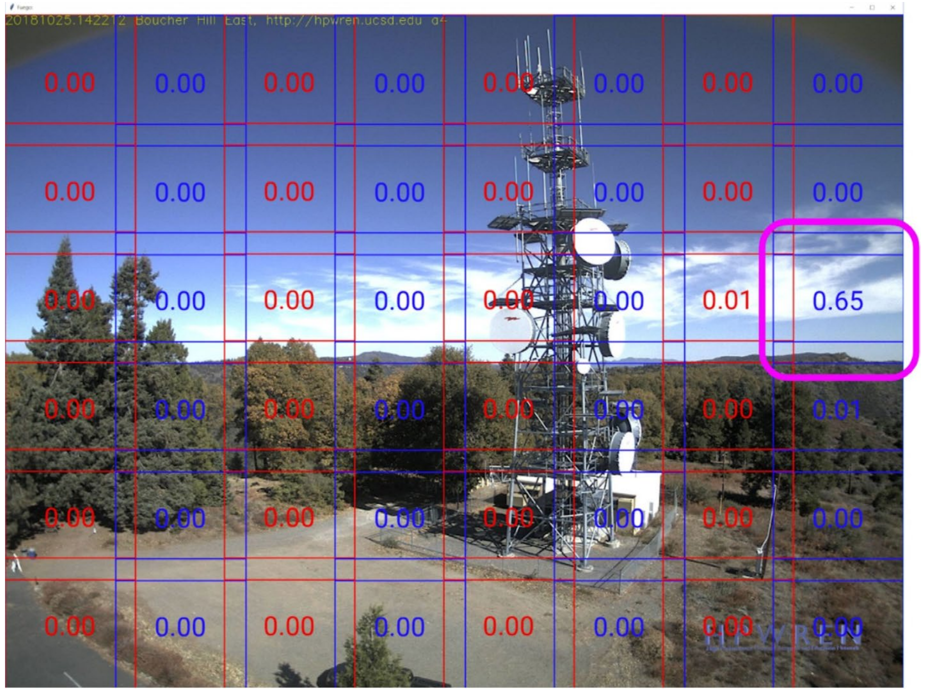
Grid it into several overlapped tiles

Image tiles:

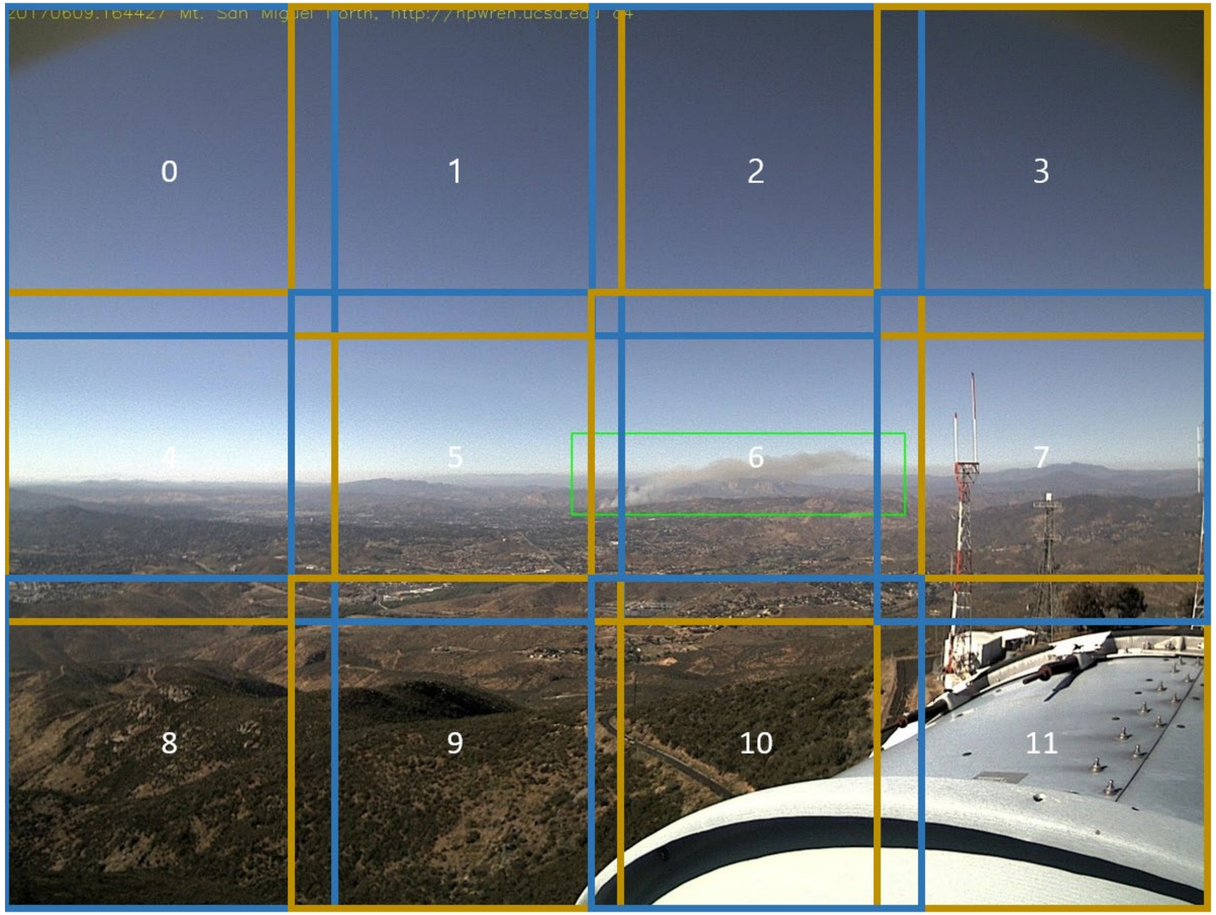
Do classification on every tile

Predict:

Get probability of smoke in every region

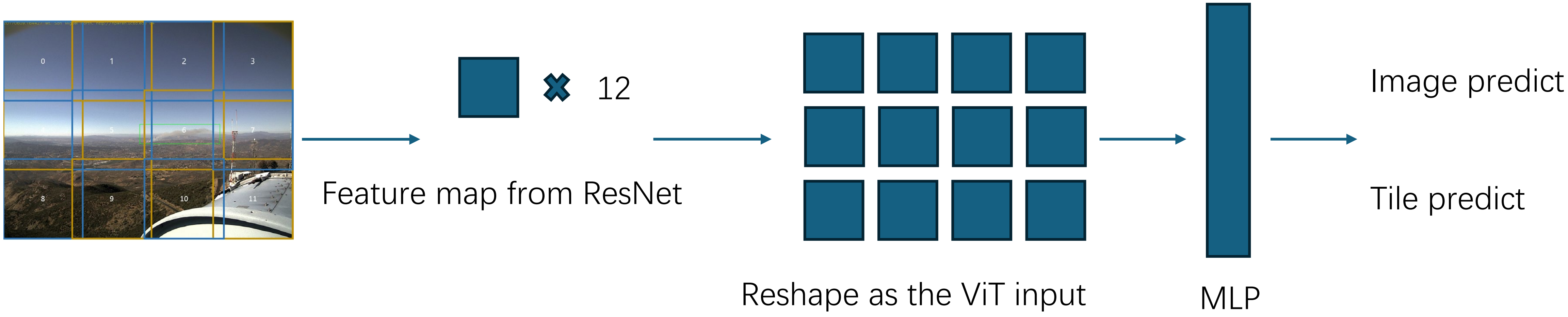


How to Define the Label



Set a threshold, then choose those tiles contains smoke area as positive(Smoke), others as negative(No smoke).

Position Encoding in ViT



Benefit of predicting on tiles:

In real implement, we are not limited to one dataset. Such kind of process enable us to deal with dynamic input and different resolutions as well as distributed process for energy management and data transmission.

A TileToImage prediction for smoke detection

Model	Paras	Time*	Acc	F1 score	Precision	Recall	TTD**
ResNet34	22.3M	50.4ms/iter	75.5	73.2	82.0	66	3.61 mins
YOLOV8	43.6M	80.3ms/iter	68.0	45.3	87	59.6	7.5 mins
Ours***	60.2M	65.2ms/iter	81	76	89	77	4.6 mins

* Time is tested on a RTX3090 in simulation for this data. And the tensorRT version can run on Jetson.
** TTD is the average detection time of smoke from the sparse sequence mapping to real time.
*** An improved SmokeyNet structure without model compression.

Results and Future Steps

- Results
  - Through our experiments, we choose a tile-to-image prediction to find the small smoke signal in the high-solution images
  - We run the TensorRT version on a Jetson AGX for energy evaluation
- Future Steps
  - We are investigating a way to not use 6-layer ViT for processing the feature data and position information which is heavy at parameter size.
  - We will optimize the TensorRT version for efficient inference.

References

- Dewangan, Anshuman, Yash Pande, Hans-Werner Braun, Frank Vernon, Ismael Perez, Ilkay Altintas, Garrison W. Cottrell, and Mai H. Nguyen. 2022. "FIgLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection" Remote Sensing 14, no. 4: 1007. https://doi.org/10.3390/rs14041007
- MMLAB @misc{mmyolo2022,title={{MMYOLO: OpenMMLab YOLO} series toolbox and benchmark},author={MMYOLO Contributors}, howpublished = {\url{https://github.com/open-mmlab/mmyolo}},year={2022}}