2024 embedded VISION SUMMIT®

10 Commandments for Building a Vision AI Product

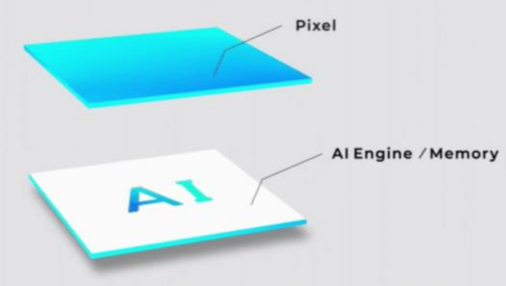
Vaibhav Ghadiok

Chief Technology Officer

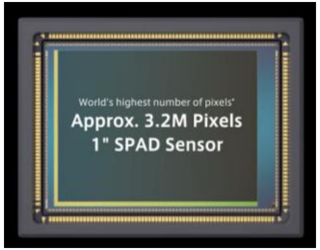
Hayden AI

HaydenAI

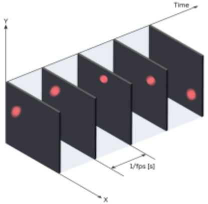
A New Era for Perception Powered by AI and Sensors



80x increase in energy efficiency of AI compute (perf/ W/ mm^2)



SPAD



Event Sensor

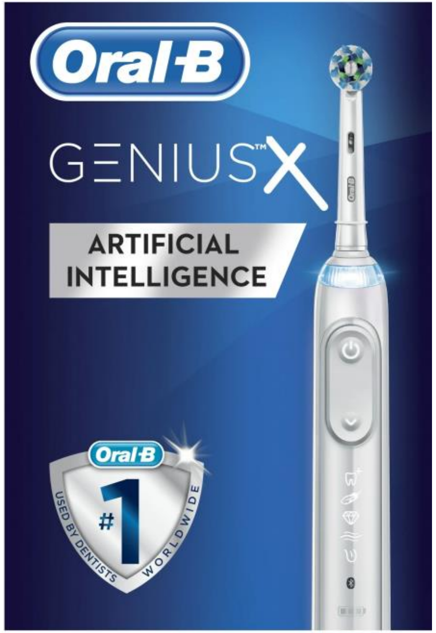
3D LiDAR

LWIR

mmWave
Radar

Thou Shalt Focus on Solving a Real-World Customer Problem

- Solve a specific problem first before solving it more generically
  - AVs in heavy trucking highways vs. urban streets
  - SLAM for ground vs. aerial vs underwater vs. AR/VR
- Separate marketing hype from actual problem solving
- Seek the simplest solution
  - The solution to everything is not AI

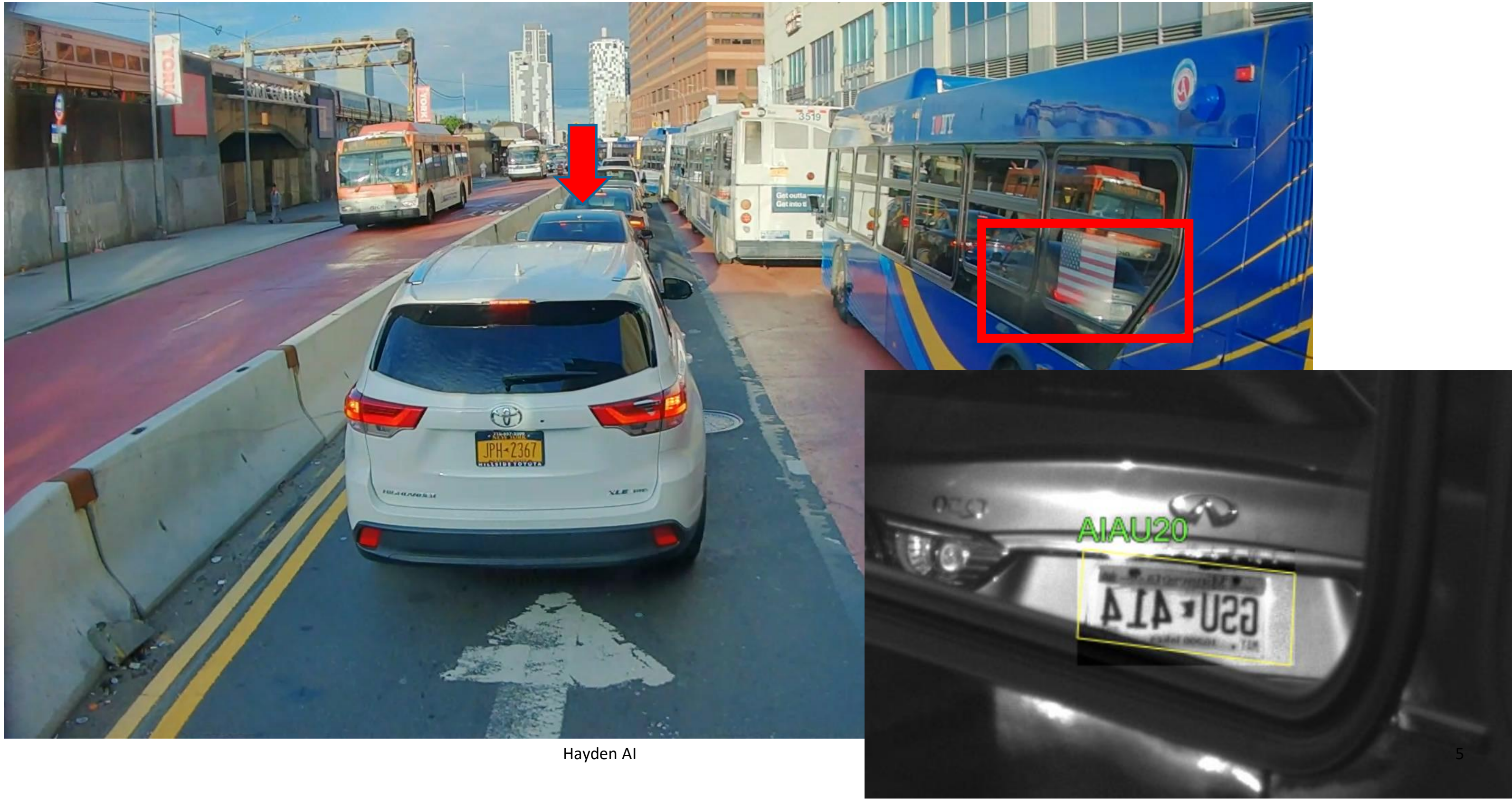


Thou Shalt Not Steal or Kill for GPUs

- Inference - How can you dramatically reduce your data/AI compute needs?

- Training - Do you require a large amount of data/compute?

Thou Shalt Not Steal or Kill for GPUs



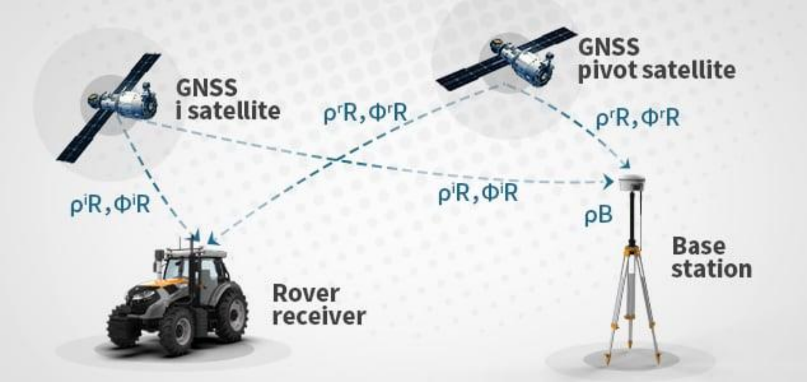
Thou Shalt Respect the Technology Gap

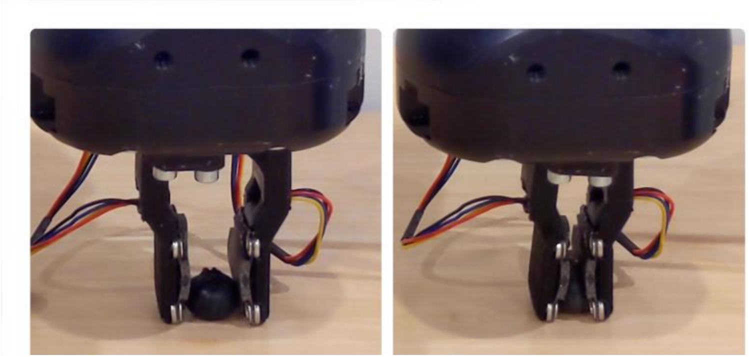
- Is there a fundamental scientific or engineering innovation needed to build the product?
  - Human-in-the loop
  - Controlling the environment

- Researcher vs. practitioner gap

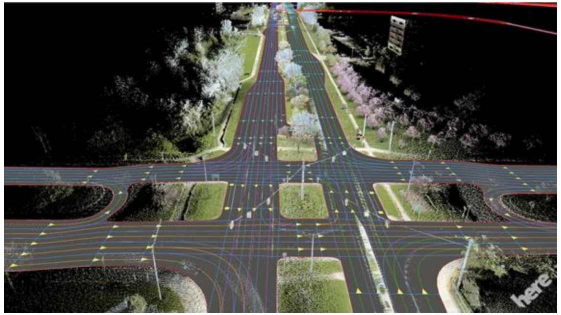
Thou Shalt Not Be a Hero

- Don’t make the problem artificially harder by limiting the sensors/actuators
  - Perception
    - Use a depth/ranging sensor
    - Real-time kinematic positioning (RTK)
    - Multispectral sensors
  - Actuator
    - Use a better gripper
    - Touch sensing with vision
    - Suction cup
  - Calibration - in-factory and in situ

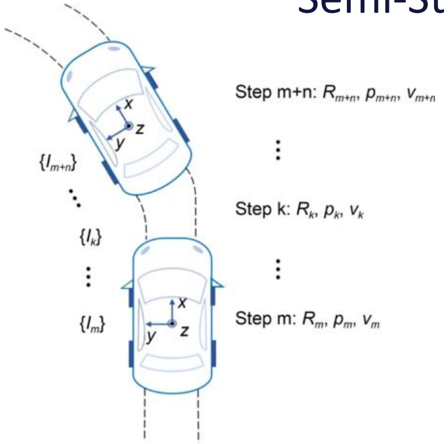


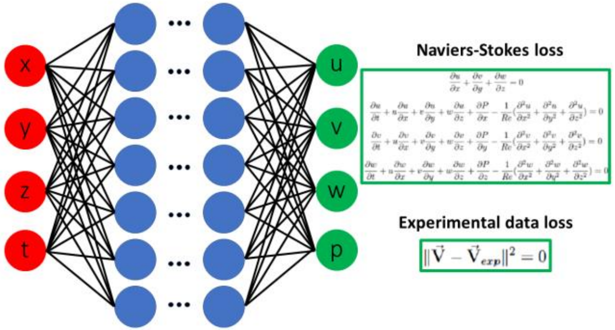


Thou Shalt Use Priors



Semi-Structured Environments

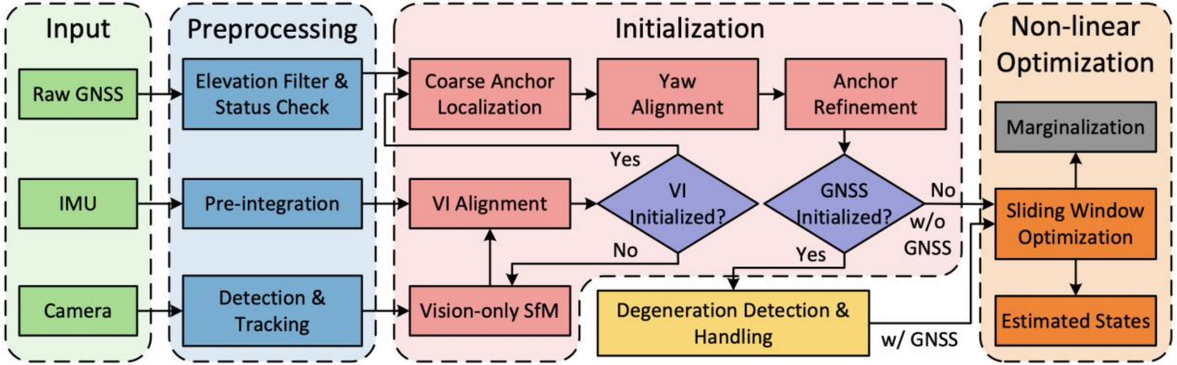

Step m+n: R_{m+n}, p_{m+n}, v_{m+n}
Step k: R_k, p_k, v_k
Step m: R_m, p_m, v_m
$\{I_{m+n}\}$
$\{I_k\}$
$\{I_m\}$
x
y
z
x
y
z


x
y
z
t
u
v
w
p
Naviers-Stokes loss
$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0$
$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}+\frac{\partial P}{\partial x}-\frac{1}{Re}(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2})=0$
$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}+\frac{\partial P}{\partial y}-\frac{1}{Re}(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}+\frac{\partial^2 v}{\partial z^2})=0$
$\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial x}+v\frac{\partial w}{\partial y}+w\frac{\partial w}{\partial z}+\frac{\partial P}{\partial z}-\frac{1}{Re}(\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+\frac{\partial^2 w}{\partial z^2})=0$
Experimental data loss
$\|\vec{\mathrm{V}}-\vec{\mathrm{V}}_{exp}\|^2=0$

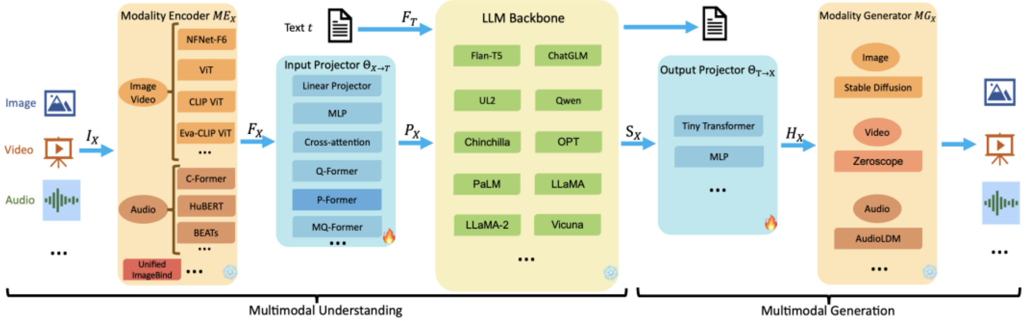
Priors

Thou Shalt Embrace Multimodality – Sensor Fusion is Good

Classical Kalman Filter / Nonlinear Optimization



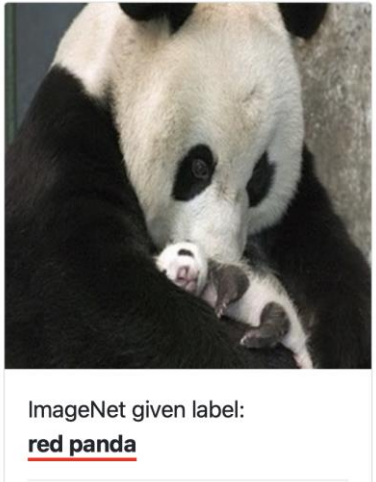
Multimodal Large Language Models



Thou Shalt Optimize for High Data Quality

- Data acquisition quality is vital
  - Carefully choose sensors - Image sensors, lens, time synchronization between sensors

- Quality of training data is critical
  - Hard negatives are good
  - AI thrives on good quality data



ImageNet given label:

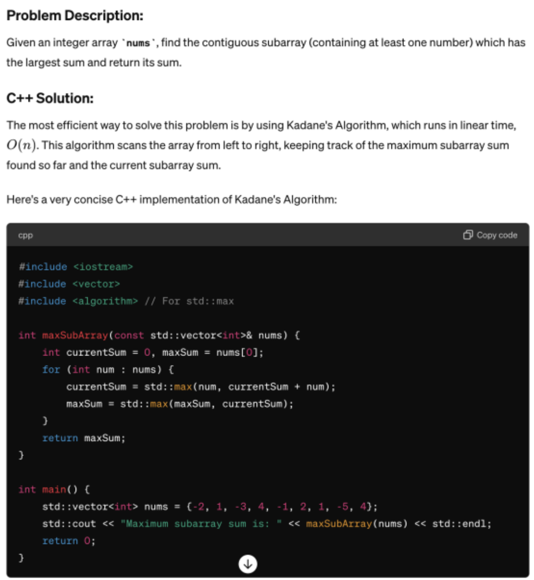
red panda

Thou Shalt Choose the Right Metrics

2023-10-04 14:40:49 UTC
LaneBus/direct | KTK2514 | 0(74) | 80f08b07-2e70-4b5e-b55a-a9f4e4fa4c53

Use more than one objective metric and ensure they are not highly correlated

Thou Shalt Not Take the Name of MM-LLMs in Vain



Problem Description:

Given an integer array `nums`, find the contiguous subarray (containing at least one number) which has the largest sum and return its sum.

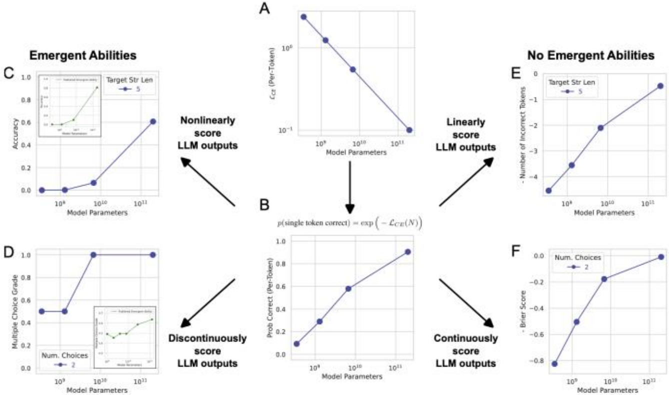
C++ Solution:

The most efficient way to solve this problem is by using Kadane's Algorithm, which runs in linear time, $O(n)$. This algorithm scans the array from left to right, keeping track of the maximum subarray sum found so far and the current subarray sum.

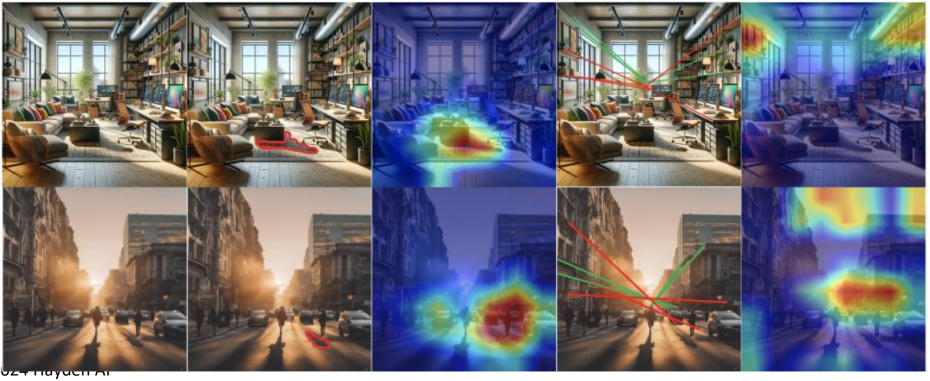
Here's a very concise C++ implementation of Kadane's Algorithm:

```cpp
#include <iostream>
#include <vector>
#include <algorithm> // For std::max

int maxSubArray(const std::vector<int>& nums) {
    int currentSum = 0, maxSum = nums[0];
    for (int num : nums) {
        currentSum = std::max(num, currentSum + num);
        maxSum = std::max(maxSum, currentSum);
    }
    return maxSum;
}

int main() {
    std::vector<int> nums = {-2, 1, -3, 4, -1, 2, 1, -5, 4};
    std::cout << "Maximum subarray sum is: " << maxSubArray(nums) << std::endl;
    return 0;
}
```

Solving a hard programming problem



No emergent capabilities



Poor in projective and Euclidean geometry

Thou Shalt Carefully Choose AI Inference Compute

- Understand total cost of development
- AI TOPS is not everything
  - Heterogenous compute
  - Memory bandwidth
  - Operator support
  - TOPS/W
  - Precision
  - Sustained compute
  - Utilization

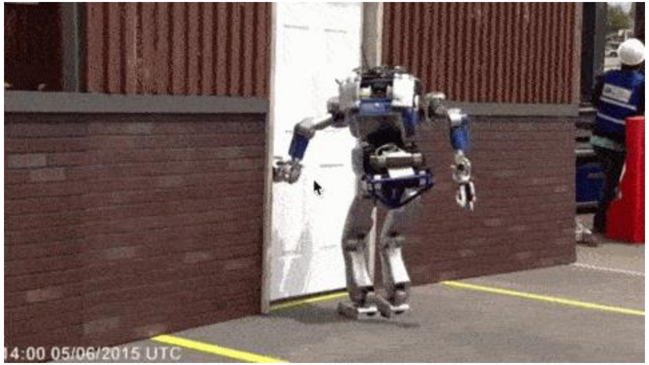
Thou Shalt Test, Continuously Learn and Adapt

- Test in the target deployment environment
  - Unlike traditional testing, 100% coverage is infeasible
- Continuously iterate and improve the system
  - Current AI is not adaptive
- Design systems to be debuggable

Conclusion

- In almost all successful deployments of AI
  - Human-in-the-loop
  - Constrain the environment
- Optimize end-to-end – multimodal sensors, AI, compute, hardware
  - The solution to every problem is not to retrain with more data
- Don’t judge AI capabilities by human analogies

vs



Resources

Are Emergent abilities of LLMs a Mirage
https://arxiv.org/abs/2304.15004

MM-LLMs – Recent Advances
https://arxiv.org/abs/2401.13601

Label Errors in ML Test Sets
https://labelerrors.com

No “Zero-Shot” Without Exponential Data
https://arxiv.org/abs/2404.04125

We are hiring!

https://www.hayden.ai/careers

Office Hours:

Wednesday, May 22, 3:30 - 4:15 pm PT
Speaker Square (across from ET-2) in the Exhibit Hall