Choose the Right Hardware

Proposal Template

Scenario 1: Manufacturing

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario? (CPU / IGPU / VPU / FPGA)
FPGA

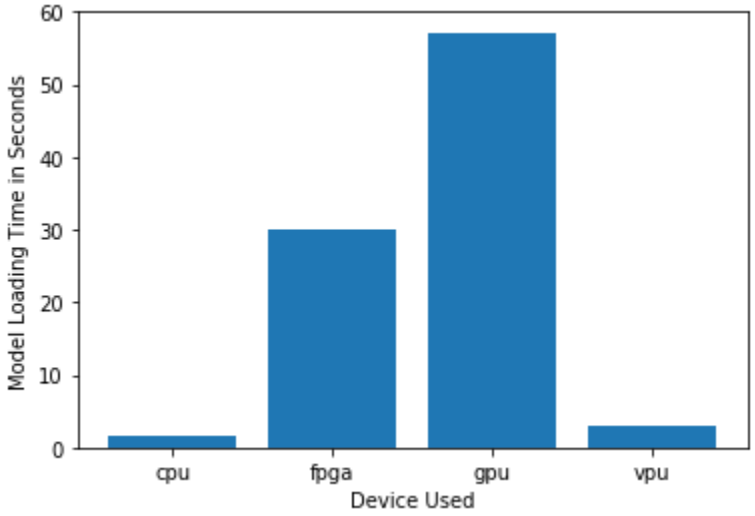
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
Client would like the image processing task to be completed five times per second.	Inference @ 5FPS is required. Most devices can provide this.
The HW must be re-purposed for another task i.e defect detection. System would need to be able to run inference on the video stream very quickly. System needs to be flexible so that it can be reprogrammed and optimized quickly for other detection tasks.	FPGA is designed for High performance and flexibility. FPGAs can also be easily repurposed to run other bitstreams for different object detection tasks without compromising on the performance.
Client has no cost constraints, but the system should last for at least 5-10 years.	FPGA are available as Embedded SKUs with long life support.

Queue Monitoring Requirements

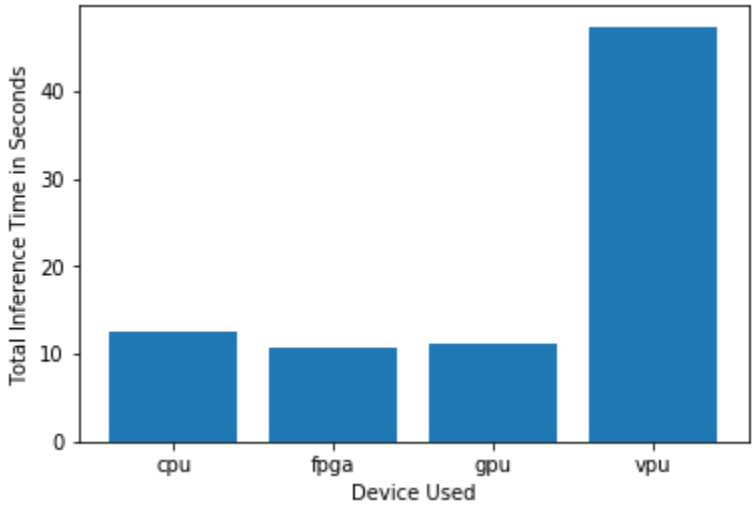
Maximum number of people in the queue	4
Model precision chosen (FP32, FP16, or Int8)	*FP16 for FPGA*

Test Results

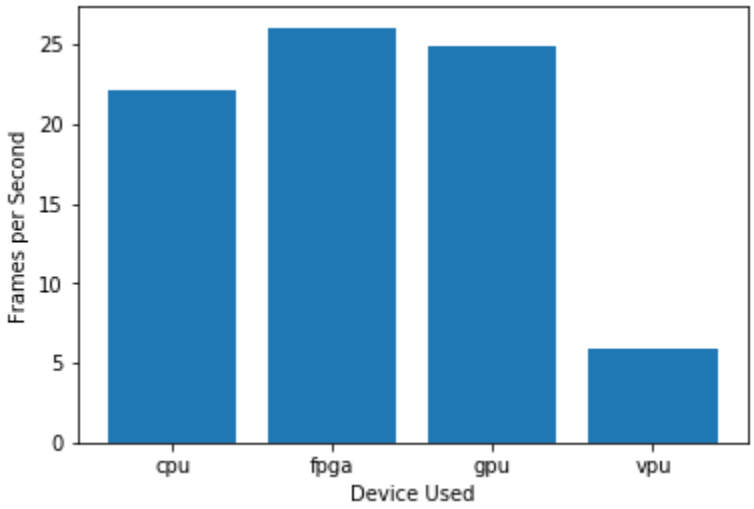
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



FPS

Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation
Camera streams are coming in at 30-35FPS and the requirement is to do inference at least 5FPS. As per the FPS chart all the devices can achieve this. *However there is also a requirement for a high-performance device with the flexibility to easily run different types of models without compromising the performance, and since cost is not a constraint,* ***FPGA*** *is a good fit for this scenario.*

Scenario 2: Retail

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario? **(CPU / IGPU / VPU / FPGA)**

CPU

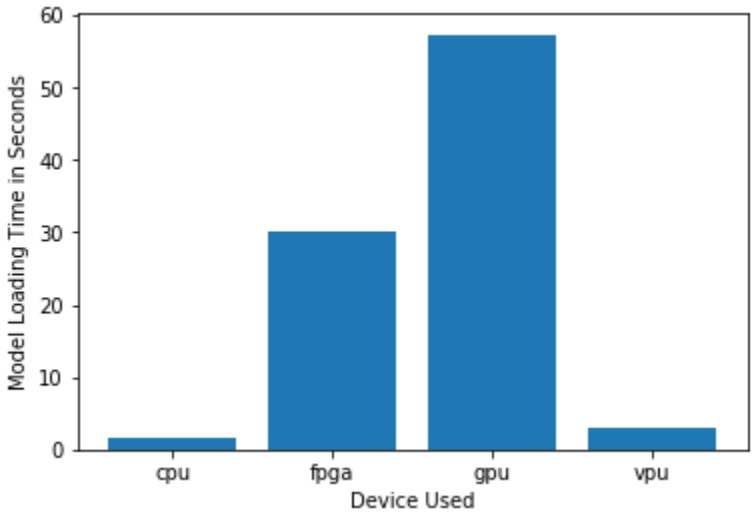
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The checkout counters already have a Core i7 and is not computationally fully loaded.	Corei7 already has a powerful CPU and GPU, an additional processor may not be required.
Client does not have much money to invest in additional hardware.	No additional HW is required since Core-i7s are already available at the checkout counter and are hardly being used.
He also would like to save as much as possible on his electric bill.	Additional HW will increase power consumption.

Queue Monitoring Requirements

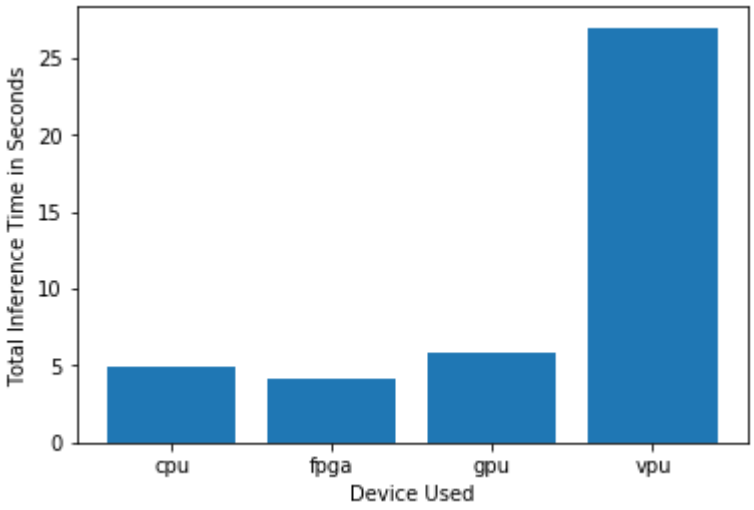
Maximum number of people in the queue	5
Model precision chosen (FP32, FP16, or Int8)	FP32 for CPU

Test Results

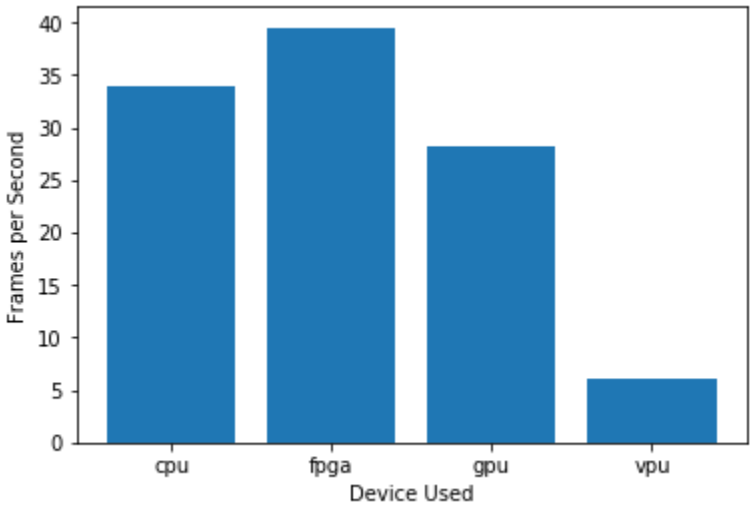
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



FPS

Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation
Adding an FPGA/VPU will incur additional costs. Moreover, the existing Core i7s at the counters, which already includes CPU and iGPU, is hardly utilized.

*It is given that customers spend 350-400secs at the checkout line during peak hours and from the FPS graph it can be noted that both CPU and GPU performs inference greater than 25 FPS (implies 25*350 frames are available to count people in the queue), which indicates either CPU/GPU satisfies the requirement.*
However, one drawback of GPU is that it has ~60secs model load time compared to ~2secs for CPU. Hence CPU is a better option.

Scenario 3: Transportation

Client Requirements and Potential Hardware Solution

Look through the scenario and find any relevant client requirements. Then, suggest a potential hardware type and explain how this hardware would satisfy each of the requirements.

Which hardware might be most appropriate for this scenario? (CPU / IGPU / VPU / FPGA)
VPU

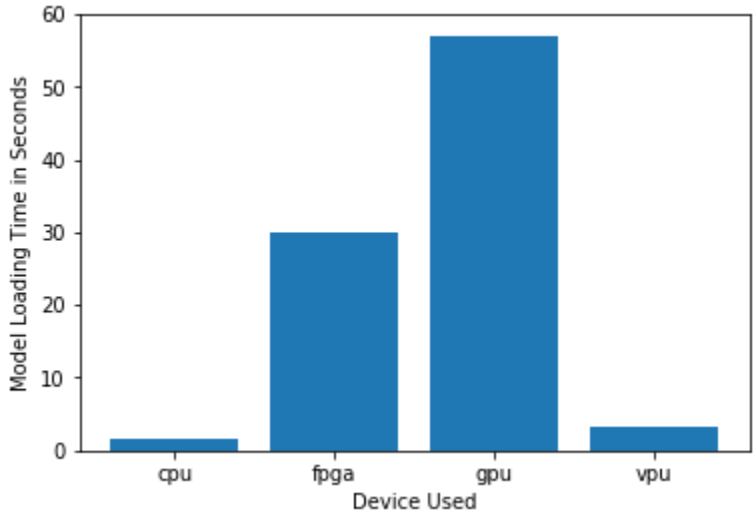
Requirement Observed (Include at least two.)	**How does the chosen hardware meet this requirement?**
No significant additional processing power is available to run inference on existing systems.	Need new HW
Budget allows for a maximum of $300 per machine.	NCS2-VPU is only $70
Client wants to save as much as possible both on hardware and future power requirements.	NCS2-VPU has only 1W TDP

Queue Monitoring Requirements

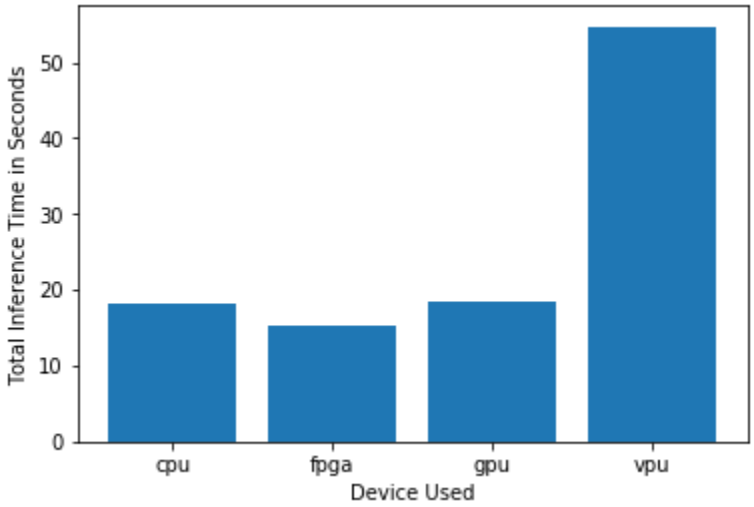
Maximum number of people in the queue	*15*
Model precision chosen (FP32, FP16, or Int8)	FP16 for VPU

Test Results

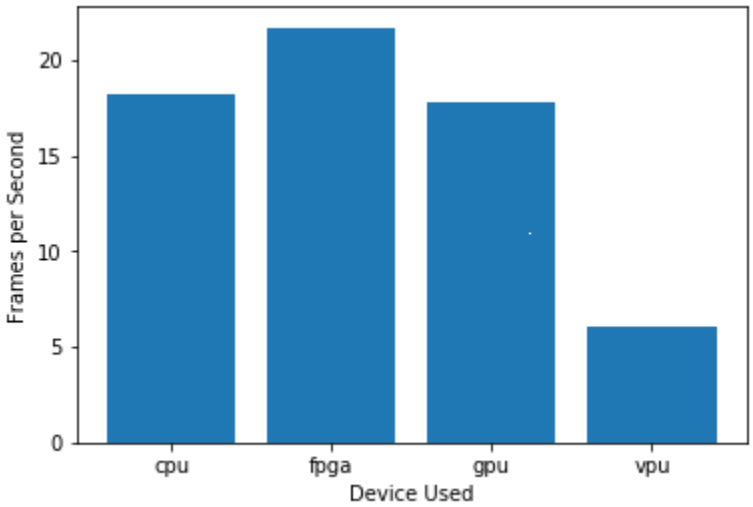
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



Model Load Time



Inference Time



FPS

Final Hardware Recommendation

Now synthesize your points from above and provide a brief write-up describing why the chosen hardware is the best choice for this scenario. Be sure to discuss the client's requirements, the test results, and how these relate to one another (e.g., perhaps one of the devices performed better than the rest, but does not meet one of the client's requirements).

Write-up: Final Hardware Recommendation
In terms of FPS, vpu is the least i.e 5 FPS, since there are 7 camera streams => 5/7 FPS *If 3 NCS2-VPUs are used => at least 2FPS per stream i.e 3*(5/7)* *During office hrs a new Train arrives every 2mins (120secs). Which means 240 (=120*2) frames across 7 cameras can be inferred. Cost of 3 NCS2 is less than $300.* *For Core i5 CPU, inference is at 18FPS and for 7 cameras => 18/7 FPS* *If 2 Corei5s are used => at least 5FPS i.e 2 *(18/7)* *However, the TDP will also be much higher i.e 35W per Corei5* *Therefore, in terms of lower budget and power requirements, VPU is the best option.*

-