Smart Queuing System

Initial Hypothesis and Recommended Hardware Proposal

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)**
In the Manufacturing scenario, the most appropriate hardware based on initial impressions would be a ***FPGA***

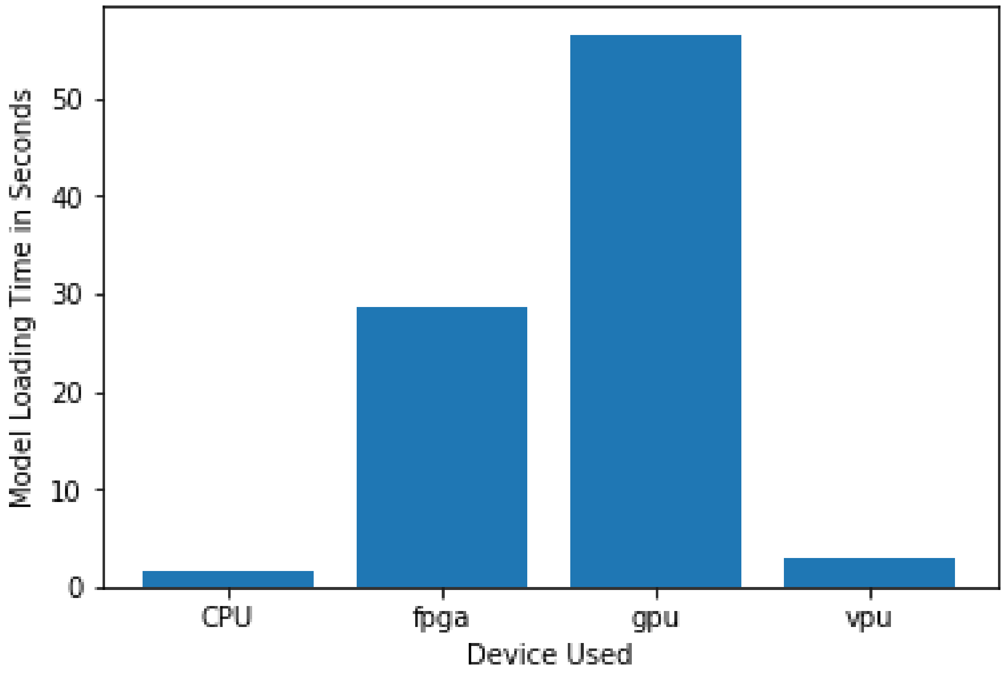
Requirement Observed **(Include at least two.)**	**How does the chosen hardware meet this requirement?**
The client has made good profits from previous sales and have plenty of revenue in hand for upgrading their infrastructure	*Because revenue to be invested here is not a problem, FPGA is a good choice based on the statement*
*The client requires a system with 24*7 up-time*	*FPGAs are the best option to ensure high intensity of working time*
The client requires a system that is flexible to changes	*FPGAs are highly flexible hardware and can be changed according to requirements*
The client requires hardware that can last for at least 5-10 years	*FPGAs have a long lifespan as compared to other hardware*

Queue Monitoring Requirements

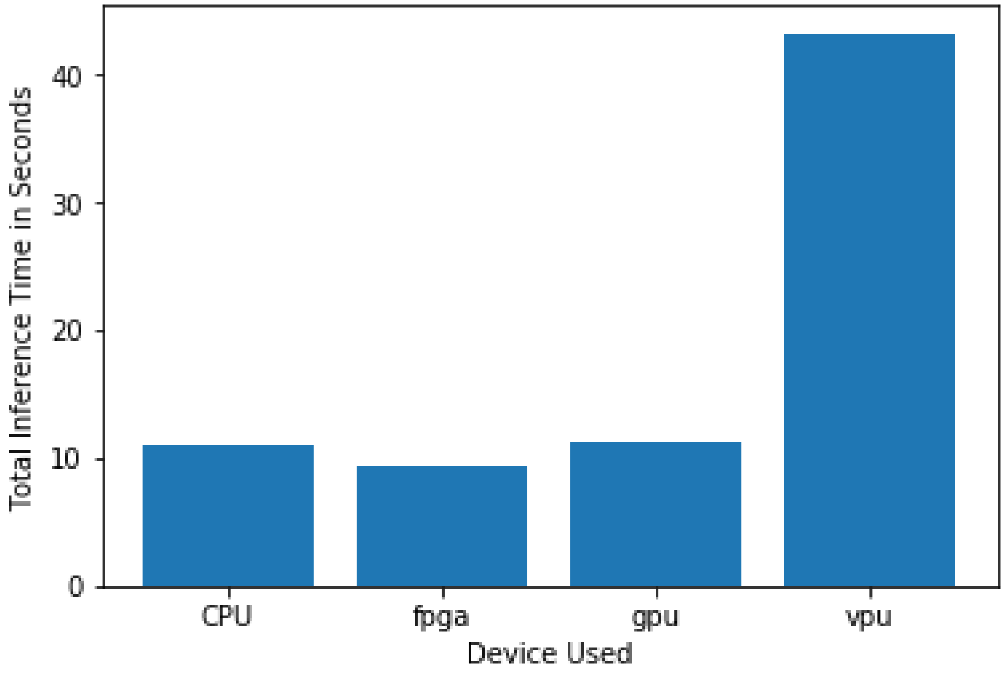
Maximum number of people in the queue	To monitor the number of people in the factory line
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

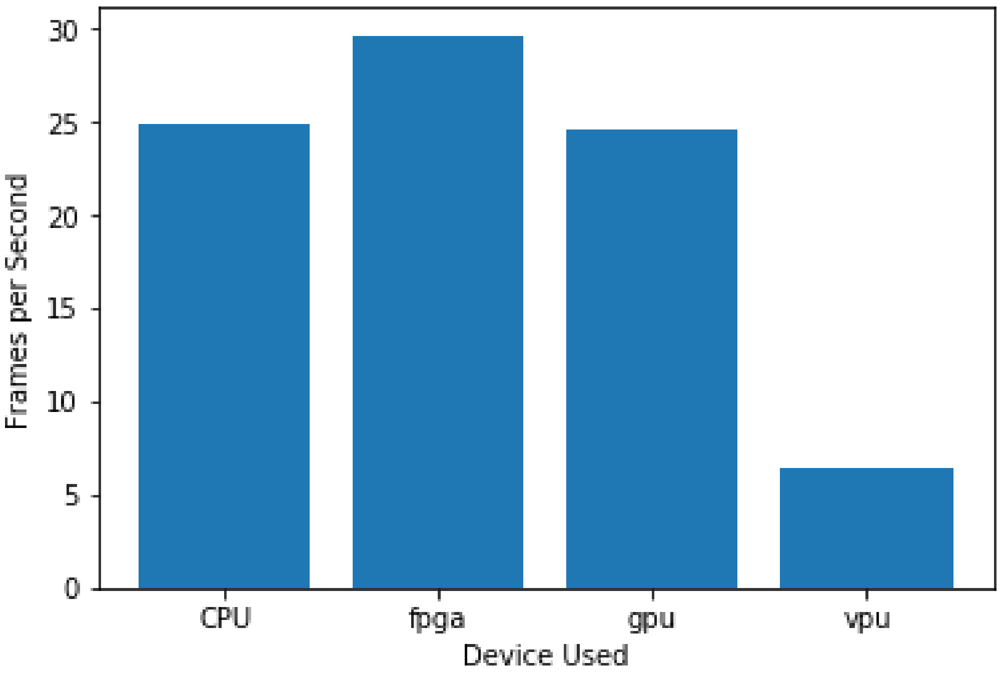
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
1. *The FPGA has higher model loading time than CPU/GPU, but the inference time is the lowest and the FPS processing is the best as we get maximum FPS reading from FPGA.* 2. *Thus, on observing the graphs, the best performance is demonstrated by the **FPGA**, as mentioned in the initial proposal and it is the recommended hardware for the client.*

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)
In the Retail scenario, the most appropriate hardware based on initial impressions would be a ***CPU***

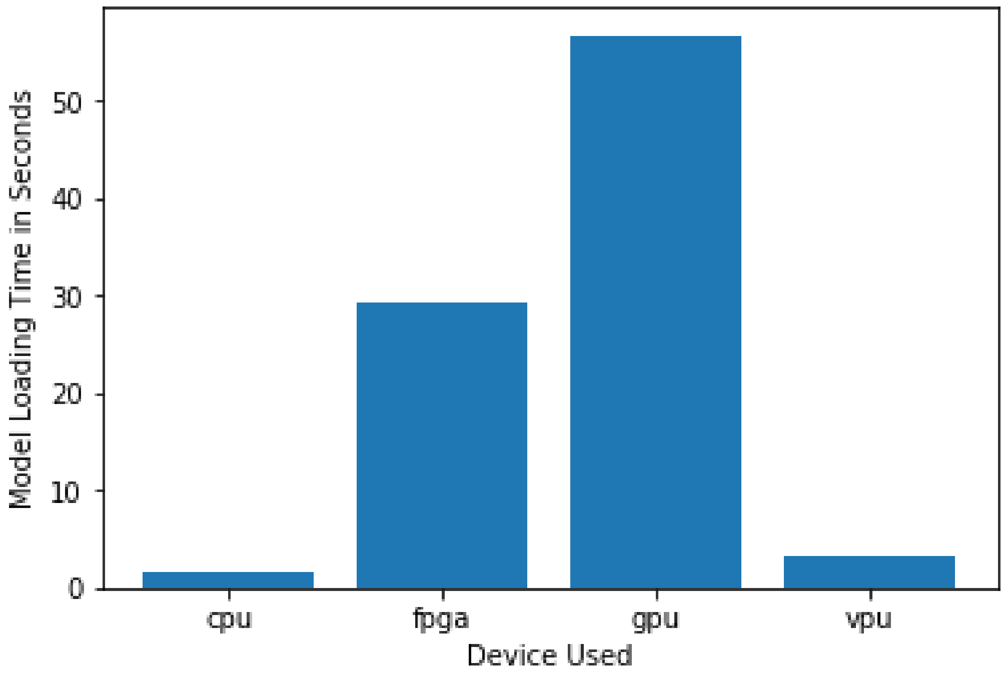
Requirement Observed (Include at least two.)	How does the chosen hardware meet this requirement?
The client store's checkout counters have modern computers with an i7 processor chip	*An i7 chip should be sufficient to run inferences using the OpenVINO toolkit*
The client does not have much money to invest in additional hardware	*Because the i7 chip is handy enough, no other hardware is required for the given use-case*
The client wants to maintain low power requirements	*CPUs have comparatively lower power consumption*

Queue Monitoring Requirements

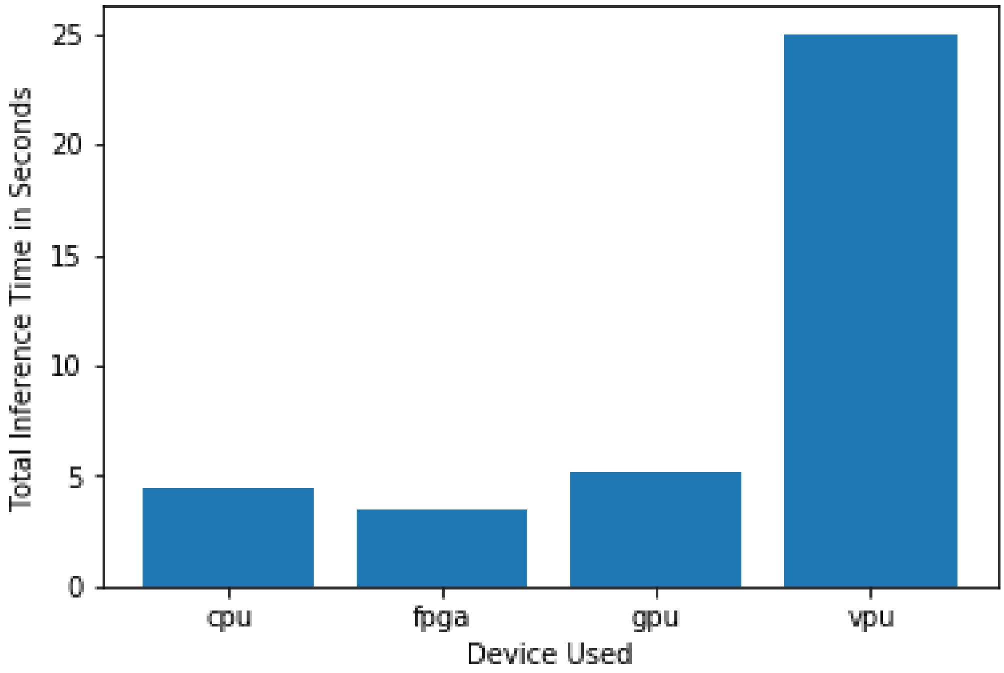
Maximum number of people in the queue	*To count the number of people in queues and direct people to the queue with lesser waiting time*
Model precision chosen (FP32, FP16, or Int8)	*FP32*

Test Results

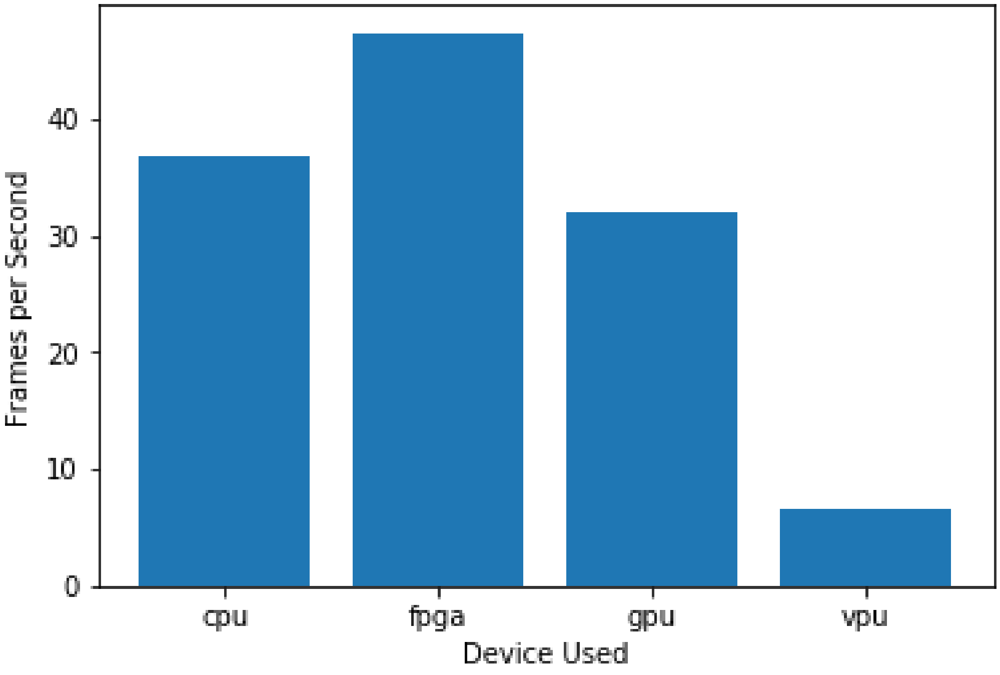
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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
1. *The model loading time for CPU is very low, which is a great indication.* 2. *The inference time for CPU is slightly higher than that of FPGA and the processing is around 10 FPS lower than FPGA.* 3. *But because the client does not have additional revenue to invest, the slight drop in performance is a fair trade-off.* 4. *Thus, based on the observations, the **CPU** is the recommended hardware for the client as mentioned in the initial proposal.*

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)**
*In the Transportation scenario, the most appropriate hardware based on initial impressions would be a **VPU***

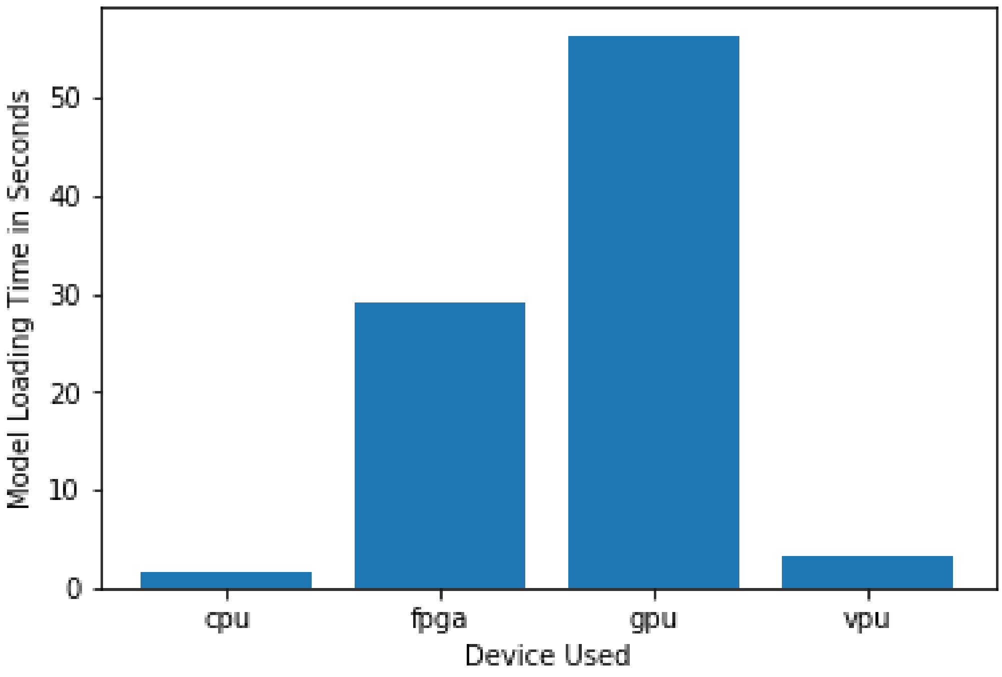
Requirement Observed **(Include at least two.)**	**How does the chosen hardware meet this requirement?**
The CPUs in these machines are currently being used to process and view CCTV footage	*The CPUs cannot be utilized for running inference*
A budget of $300 per machine is allocated for upgrading hardware	*The VPU's cost is covered within the stipulated budget*
The client wants to save on power and future requirements	*VPUs consume low power, so it is an ideal choice in this case*

Queue Monitoring Requirements

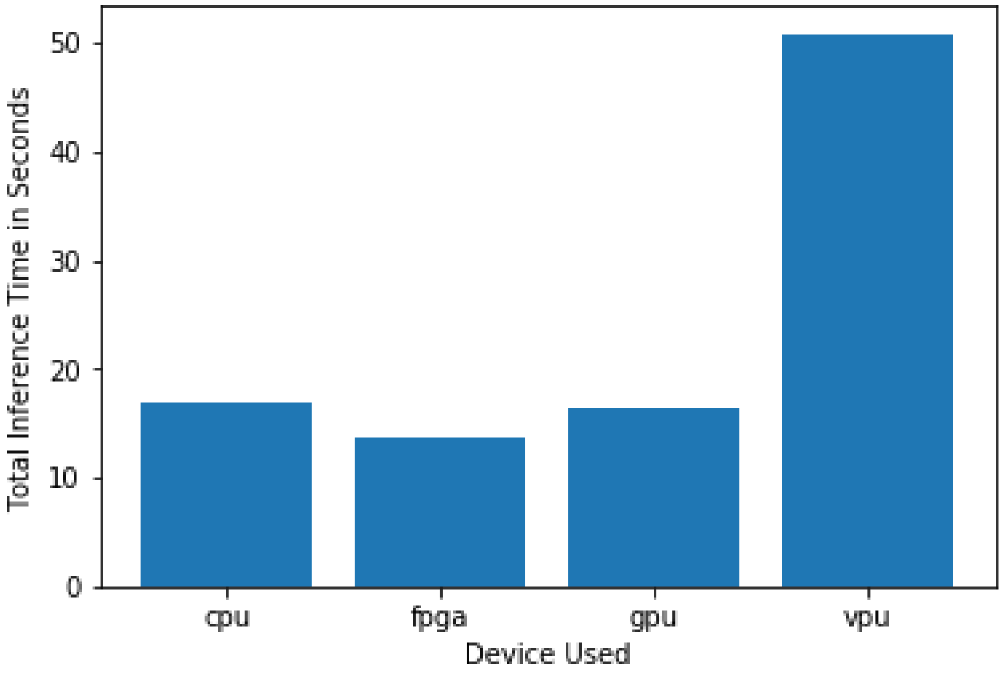
Maximum number of people in the queue	*Count the number of people in queues and help redirect crowd to decongest the metro station*
Model precision chosen (FP32, FP16, or Int8)	*FP16*

Test Results

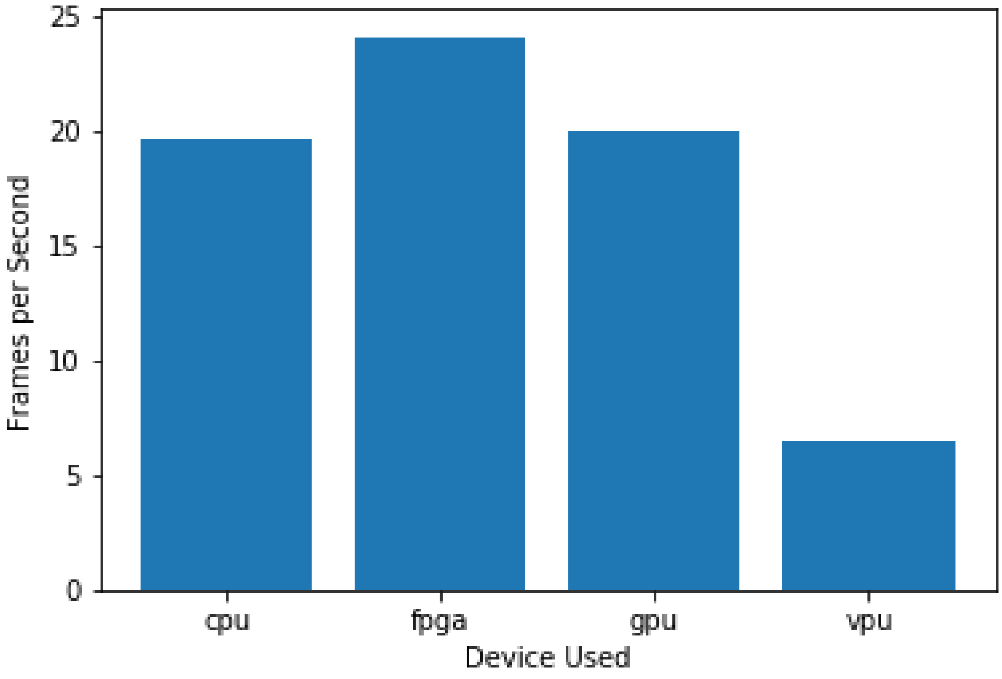
After you've tested your application on all four hardware types (CPU, IGPU, VPU, and FPGA), copy the matplotlib output showing the comparison into the spaces below. You should have three graphs (for model load time, inference time, and FPS).



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

1. *The budget mentioned for upgrades is 300$, but the client intends to save as much as possible on cost and power requirements.*
2. *The best candidate as mentioned in the initial proposal was the **VPU**. However, if we observe the graphs, we can observe that though the model loading time for the VPU is significantly lower, the inference time for the VPU is very high, which affects the processing rate of the frames during inference.*
3. *Also, the inference time for the given use-case should be as low as possible, because the train boarding time is as low as a few seconds. Thus, the FPS rate should be high and inference time of hardware should be low.*
4. *Thus, even though the GPU (Intel® Core™ i5-6500TE processor with Intel® HD Graphics 530 integrated GPU) setup costs ~1.8 times that of the VPU (192$ for the CPU+GPU vs 110$ for the VPU), the extra investment is worth it for the purpose. Also, considerations of batch processing in the application can also be made, which can further enhance performance.*
5. *Considering the graphs and points made as mentioned above, the final recommended hardware is the **CPU+GPU** setup*