Choose the Right Hardware

Proposal by Rob Straker
April 30, 2020

Scenario 1: Manufacturing

Which hardware?

Which hardware is most appropriate for this scenario? (CPU + Intel GPU / CPU + FPGA / CPU + VPU)
CPU + FPGA

Requirements

Now that you've picked the hardware, it's time to explain *why* this hardware is the right choice. Look through the scenario and find any relevant requirements. Be sure that you at least include the following:

1. Power Requirements
2. Space Requirements
3. Economic Constraints

Describe each requirement below, along with an explanation of how the selected hardware meets that requirement.

Requirement observed	How does the chosen hardware meet this requirement?
Power Constraints *- The client is moving towards an energy efficient workplace.*	*FPGAs can be more power-hungry, but the Intel Arria 10 consumes just 30 W, which is likely a fraction of what the system power supply would consume, and negligible in comparison with other power costs like lighting and heating.*
Space Constraints *- There are no stated space constraints, but it is likely that it is preferable that devices be added to existing computers.*	*An FPGA card would fit in a PCIe slot within the current chassis of the computers already present.*
Economic Constraints *- There are no stated constraints related to costs, so cost does not appear to be much of a factor.*	*Although FPGAs tend to be among the more costly devices, the cost is less than for ASIC in the long run.*
Flexibility Constraints *- The client wants a device that has future flexibility so they can reprogram and optimize the system for different use-cases.*	*FPGAs are flexible in a few different ways. They are field-programmable; they can be reprogrammed to adapt to new, evolving, and custom networks.*

Environment Constraints *- The client wants a system that can operate in a manufacturing environment.*	*FPGAs are designed to have 100% on-time performance, meaning they can be continuously running 24 hours a day, 7 days a week, 365 days a year. They are also able to function over a wide range of temperatures, from 0° C to 60° C. This means that FPGAs can be deployed in harsh environments like factory floors and still perform optimally.*
Lifespan Constraints *- Ideally, the client would like the system to last at least 5-10 years.*	*FPGAs have a long lifespan. For example, FPGAs that use devices from Intel's IoT Group have a guaranteed availability of 10 years from the start of production.*
Performance Constraints *- The client would like the system to run inference on the video stream very quickly, so it can detect chip flaws without slowing packaging.*	*FPGAs are generally the highest performing type of accelerator. Once programmed with a suitable bitstream, FPGAs can execute neural networks with high performance and very little latency. Experimentation showed.* *Experiments on Intel DevCloud edge devices showed that the FPGA had the lowest Average Inference Time and highest Frames per Second. Although Model Loading time was the longest, this need only be done once at system startup, so it doesn't impact frame processing performance.*

Write-up

Now synthesize your points from above and provide a brief write-up (not more than about 50 words) describing why the chosen hardware is the best choice for this scenario.

Write-up: Why is this the right hardware?
An FPGA device would make the most sense for the client. A removable NCS2 device would not be the best option, as it is to be installed on the factory floor, which is an open environment. And a CPU or GPU would not provide the flexibility, robustness, lifespan, or performance desired. This leaves an FPGA card which would fit in a PCIe slot within the chassis of existing computers.

Queue Monitoring Requirements

Number of people required per queue:	*2 per packaging queue (to match average during shift)*
Time for the process in the queue (ms/s):	*200ms/s (to process images 5 times/second)*
Model precision chosen (FP32, FP16, or Int8)	*FP16 or FP32 (precisions available for IEI Mustang-F100-A10)*

Scenario 2: Retail

Which hardware?

Which hardware is most appropriate for this scenario? (CPU+In GPU / CPU+ FPGA / CPU+VPU)
CPU + Integrated GPU

Requirements

Now that you've picked the hardware, it's time to explain *why* this hardware is the right choice. Look through the scenario and find any relevant requirements. Be sure that you at least include the following:

1. Power Requirements
2. Space Requirements
3. Economic Constraints

Describe each requirement below, along with an explanation of how the selected hardware meets that requirement.

Requirement observed	How does the chosen hardware meet this requirement?
Power Constraints *- The client wants to save as much as possible on his electric bill.*	*The CPU + Integrated GPU already exists (Intel i7 core processor), so there are no additional power requirements. A VPU or FPGA solution would require an additional device which would require additional electrical power.*
Space Constraints *- The client does not have any store floor space available, and currently already has computers at each checkout counter.*	*The CPU + Integrated GPU is present in existing computer systems (Intel i7 core processor), so additional space would not be required.*
Economic Constraints *- The client does not have much money to invest in additional hardware.*	*The CPU + Integrated GPU is present in existing computer systems for most checkout counters (Intel i7 core processor), so additional devices would need to be purchased for only a few counters without them.*
Environment Constraints *- The client wants a system that can operate in a retail environment.*	*A CPU + Integrated GPU is sufficiently robust to withstand the temperature range and cleanliness of a typical retail environment.*
Lifespan Constraints *- The client has not expressed any requirements for lifespan, but presumably they would like a system that will not need to be replaced or upgraded for several years at a minimum.*	*A modern CPU + Integrated GPU like the Intel i7 core processor is relatively long-lasting.*

Performance Constraints *- The client has not specified particular requirements for performance, but the system will need to be able to monitor queues that form during waits at counters during the busiest times during the day.*	*A modern CPU + Integrated GPU like the Intel i7 core processor, currently used to carry out some minimal tasks that are not computationally expensive, would provide sufficiently high performance.*

Write-up

Now synthesize your points from above and provide a brief write-up (not more than about 50 words) describing why the chosen hardware is the best choice for this scenario.

Write-up: Why is this the right hardware?
A CPU + Integrated GPU device would make the most sense for the client. The client wants to save on electricity costs, does not have any store floor space available, and does not have much money to invest. However, he has personal computers installed with Intel i7 core processors for most checkout counters, so CPU + Integrated GPUs are already available.

Queue Monitoring Requirements

Number of people required per queue:	*2 per counter (to match average in normal daily hours)*
Time for the process in the queue (ms/s):	*200ms/s (for real-time monitoring during busiest periods)*
Model precision chosen (FP32, FP16, or Int8)	*FP16 (IGPUs are optimized for FP16)*

Scenario 3: Transportation

Which hardware?

Which hardware is most appropriate for this scenario? (CPU+In GPU / CPU+ FPGA / CPU+VPU)
CPU + VPU

Requirements

Now that you've picked the hardware, it's time to explain *why* this hardware is the right choice. Look through the scenario and find any relevant requirements. Be sure that you at least include the following:

1. Power Requirements
2. Space Requirements
3. Economic Constraints

Describe each requirement below, along with an explanation of how the selected hardware meets that requirement.

Requirement observed	How does the chosen hardware meet this requirement?
Power Constraints *- The client wants to save as much as possible on future power requirements.*	*VPUs are low-power devices. While the TDP of a CPU ranges from 40-100 watts, a VPU can reduce this by a factor up to 8 times. For example, the Myriad X has a power consumption of only 1-2 watts.*
Space Constraints *- The client already has 7 CCTV cameras on the platform connected to PCs located in a nearby security booth, so there is not a lot of available space for new devices.*	*A VPU such as the NCS2 comes in a small form factor, 72.5mm X 27mm X 14mm, and looks like a standard thumb drive. It has a USB interface (2.0, 3.1), so it can be plugged directly into existing systems.*
Economic Constraints *- The client has a maximum budget of $300 per machine, and would like to save as much as possible on hardware requirements.*	*VPUs provide a cost-efficient way to add performance to a pre-existing system. Compared to other AI accelerators, the NCS2 is an inexpensive option, typically costing around $70 to $100.*
Environment Constraints *- The client wants a system that can operate in a busy urban passenger transportation environment.*	*A VPU would be plugged into existing systems that are located in a security booth, so it would be secure from accidental removal or theft. It would also be protected by the extremes of temperature and dirt. In any case, it can operate in a 0-40 Celsius environment.*
Lifespan Constraints *- The client has not expressed any requirements for lifespan, but presumably they*	*All of Intel's devices are relatively long-lasting.*

would like a system that will not need to be replaced or upgraded for several years at a minimum.	
***Performance Constraints** - The client has not specified particular requirements for performance, but the system will need to be able to monitor queues that form at train doors during the busiest times during the day.*	*This is the lowest performing type of accelerator. They can be used to accelerate the performance of a pre-existing system. They have low latency due to on-chip memory.*

Write-up

Now synthesize your points from above and provide a brief write-up (not more than about 50 words) describing why the chosen hardware is the best choice for this scenario.

Write-up: Why is this the right hardware?
A CPU + GPU option is not suitable, since Ms. Leah's current PC system is used for CCTV recording, so there may not be enough capacity in this machine to perform inference. Therefore, some kind of add-in accelerator card is recommended. An FPGA card would cost over $1,000, which exceeds Ms. Leah's budget of $100 to $150. Therefore, one or two NCS2 sticks would be the best option.

Queue Monitoring Requirements

Number of people required per queue:	*7 per door (to match average in non-peak hours)*
Time for the process in the queue (ms/s):	*200ms/s (for real-time monitoring during busiest periods)*
Model precision chosen (FP32, FP16, or Int8)	*FP16 (NCS2 only supports FP16 model precision)*