Intel Edge_Computing Resume

Chapter I:

What is the Edge ?

- Local processing.
- Not just anywhere in the cloud.
- Edges Applications are often used where low latency is necessary.
- Used where a network may not always be available.
- Can come from a desire for real time decision making.

-When using the Edge ,there is no need to send data to the cloud =>

-**More secure** (protect from DoS attack « Denial of Service »

-**Less impact** on the network.

The Edge Computing doesn't necessarily mean no cloud ,Edge AI Algorithm for example can still be trained in the cloud but get **deployed at the Edge.**

Why is AI at the Edge Important ?

-Network impact : communication can be expensive in term of:

-**bandwith.**

-**power consumption.**

Sometime network communication can be impossible ,even if UDP packets are send in the case of video and audio streaming ,it'll still be difficult to get the right information .

-Latency considerations: Imagine a self driving car which react too slowly ,it would be so bad !

Security: Personal / Health data if sent to the cloud could be used for bad purpose or simply for marketing ,even Business Data could be stolen.

-Optimization software :can help achieve great efficiency with edge AI models.

The 3 reasons for development of the Edge:

-Proliferation of devices.

-Need for low-latency compute.

-Need for disconnected devices.

What is a Pre-Trained Model ?:

The pre-trained model is generally used in ***Transfer Learning*** . It is popular in computer vision because it allows us to build accurate models in a timesaving way.

Instead of starting the learning process from scratch you start from patterns that have been learned when solving a different problem.

Example of Pre-trained Model : *VGG* , *ResNet ,AlexNet*

The Inference engine: *Perform inference on an Intermediate representation (R format)*

Open Model Zoo help us obtain a pre-trained model . If you have your model ,you can convert it with the Model Optimizer ,and the perform inference on it with the Inference Engine.

If you created your own model, or are leveraging a model not already in IR format (TensorFlow, PyTorch, Caffe, MXNet, etc), use the ***Model Optimizer*** first. This will then feed to the Inference Engine, which can be integrated into your app and deployed at the edge, so it converts a model for further use with the toolkit.