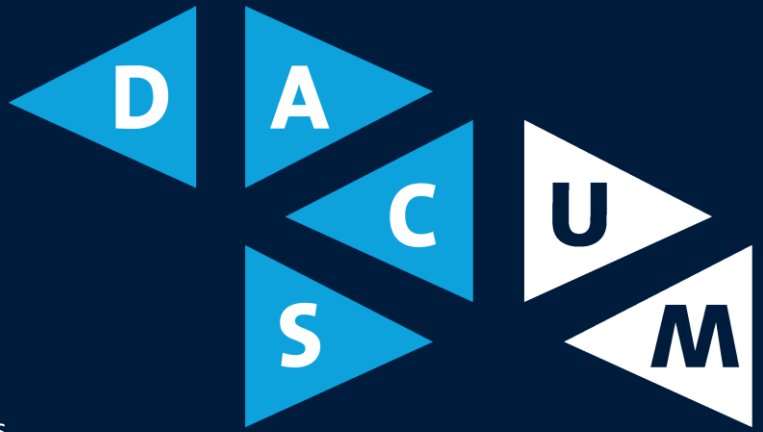
A segmented robot grasping perception neural network for edge AI
IEEE SMC 25

Casper Bröcheler, Thomas Vroom, Derrick Timmermans,
Alan van den Akker, Guangzhi Tang,
Charis Kouzinopoulos, Rico Möckel

Presenter: Dr. Charis Kouzinopoulos
Computer Systems research group



Maastricht University | Department of Advanced Computing Sciences

Object grasping (in cluttered scenes) is a core function of robotics

identification

reaching

grasping

Steps involved

CoRoSect Horizon Europe - Insect Breeding

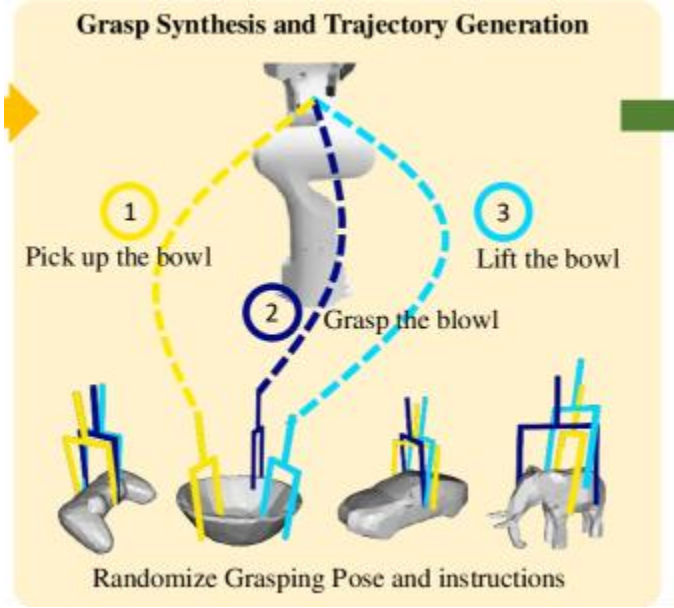
Patient handling

Assembly lines

Object sorting

Application areas

- **Grasp Synthesis:** The problem of creating a set of grasp poses
- Robotic manipulators can grasp objects from multiple angles. 6-DOF *grasp pose detection* enables robots to grasp from any angle!
- Fast and accurate grasping is still challenging for robots. Tend to generate unreliable grasps, especially in **cluttered scenes**
- Most works resort to the **whole point cloud** for grasp generation
- **HGGD** is a very promising SoTA model that aggregates instead multiple graspable local areas - though computational expensive!



S. Deng et. al., GraspVLA: a Grasping Foundation Model Pre-trained on Billion-scale Synthetic Action Data

Research questions

RQ1: "How to execute HGGD at the edge with minimal energy consumption for localized decision making?"

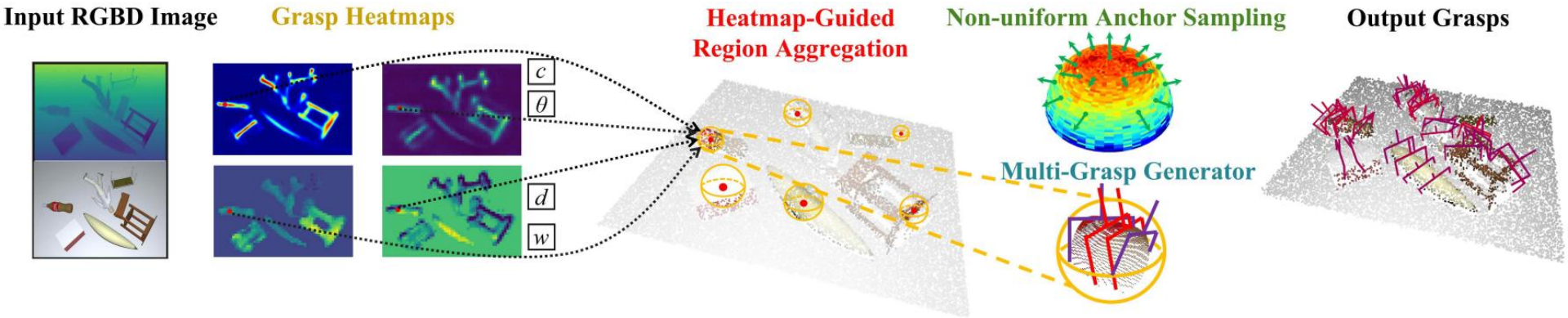
RQ2: "How to parallelize HGGD to enable faster execution at the edge?"

Part A: The HGGD model

Combination of RGBD images
with grasp heatmaps

Position Orientation Opening width

Siang Chen, et. al., Efficient Heatmap-Guided 6-Dof Grasp Detection in Cluttered Scenes. 2024

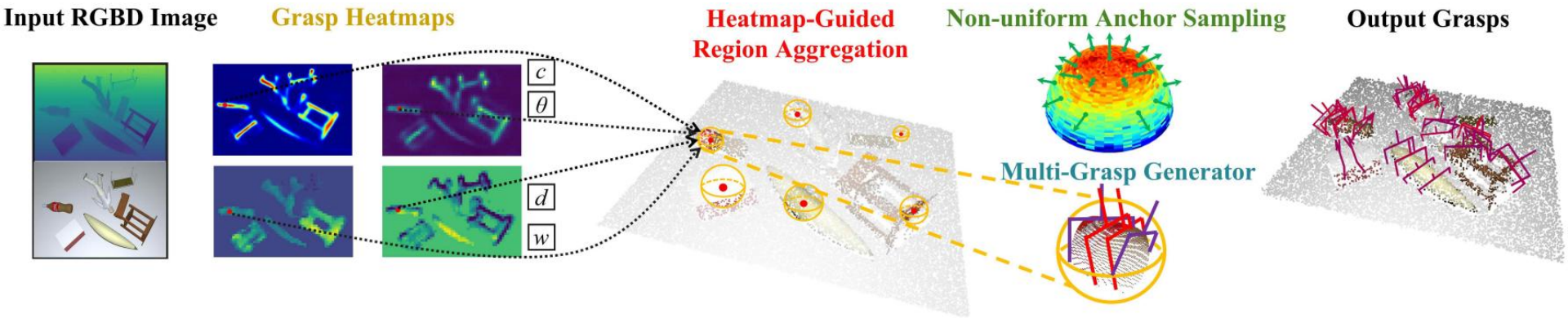


Combination of RGBD images with grasp heatmaps

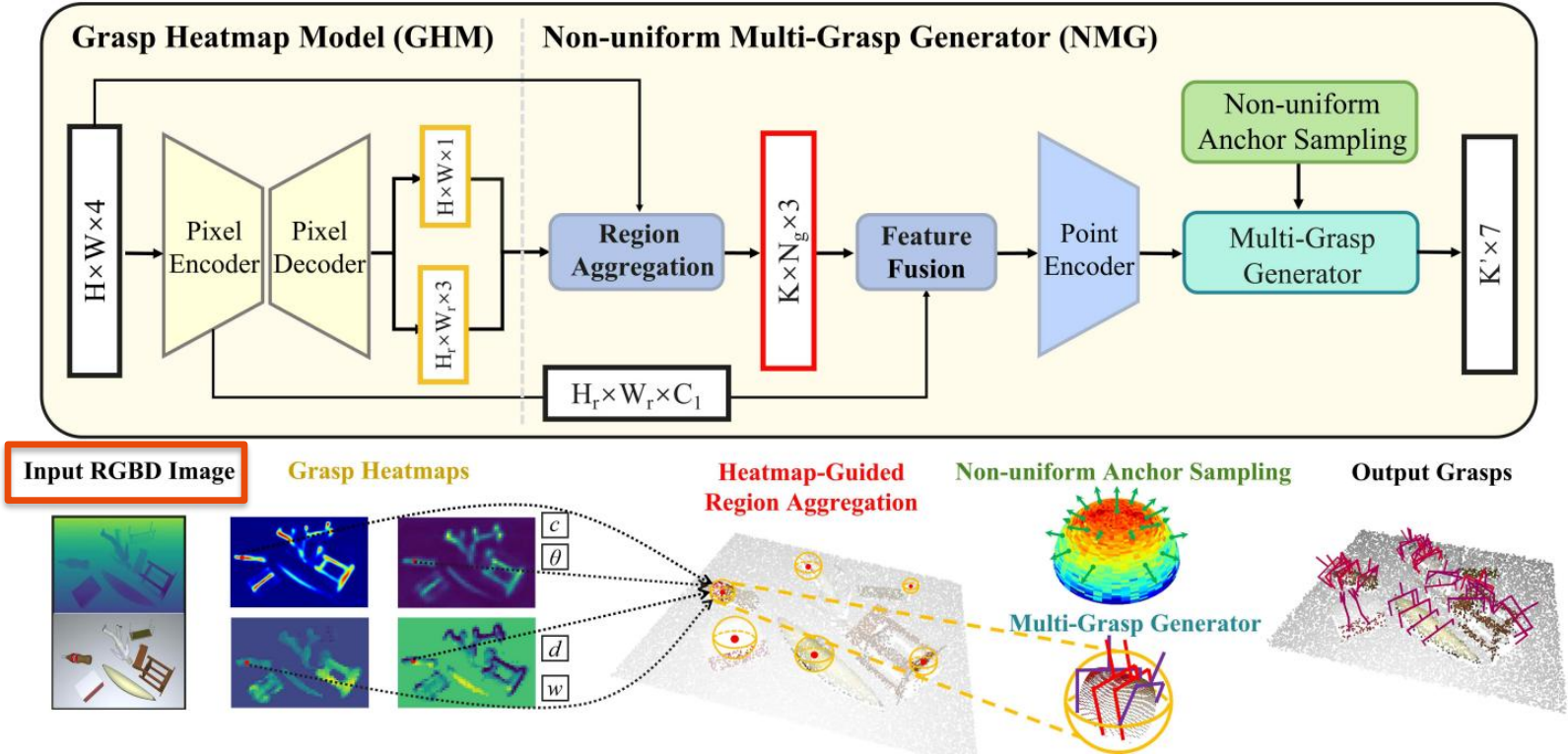
Grasp heatmaps can be generated as guidance to **aggregate local points** into **graspable regions** for further grasp pose generation

Position Orientation Opening width

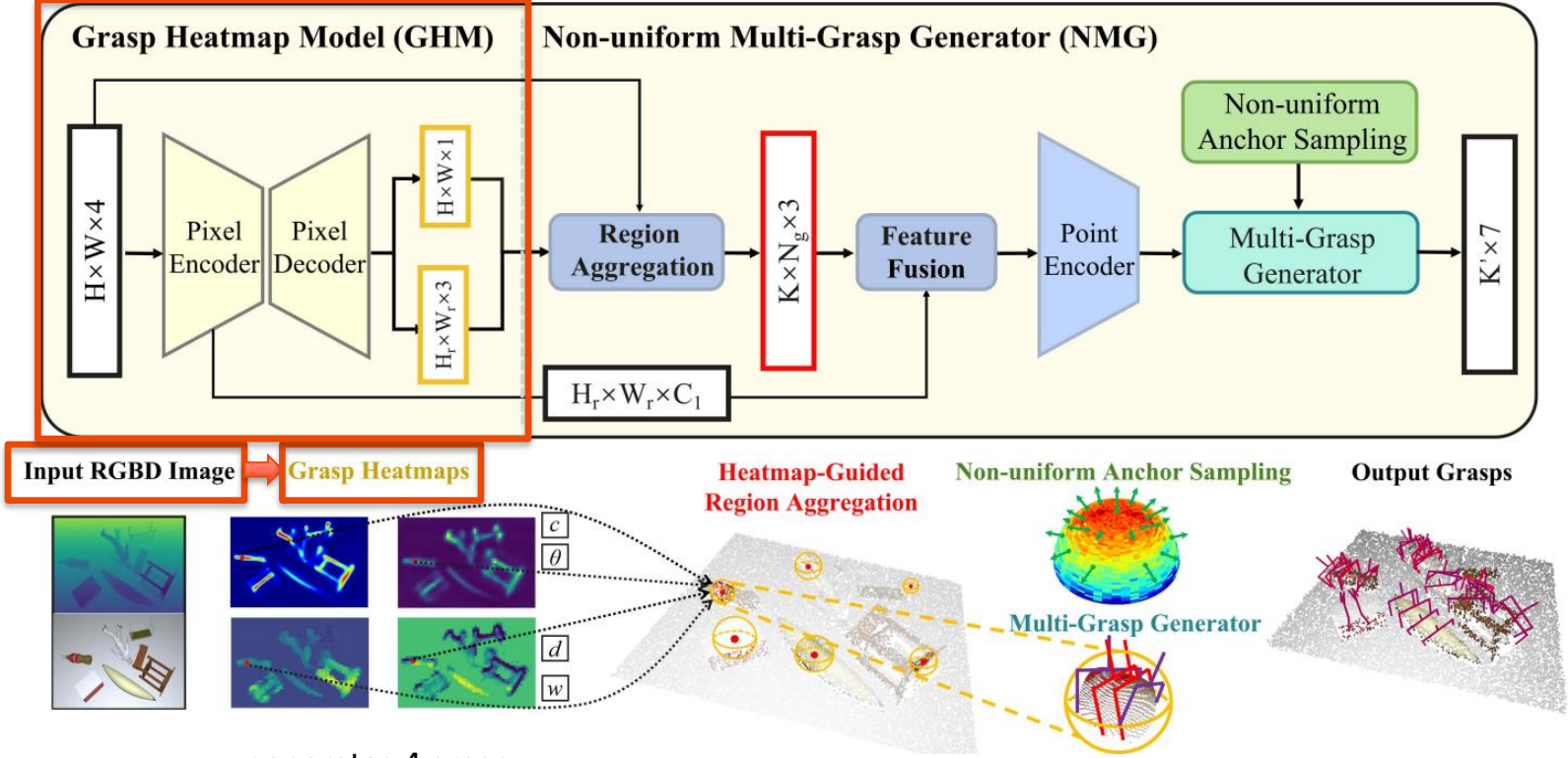
Siang Chen, et. al., Efficient Heatmap-Guided 6-Dof Grasp Detection in Cluttered Scenes. 2024



Encoder-decoder - extracts semantic features via a CNN

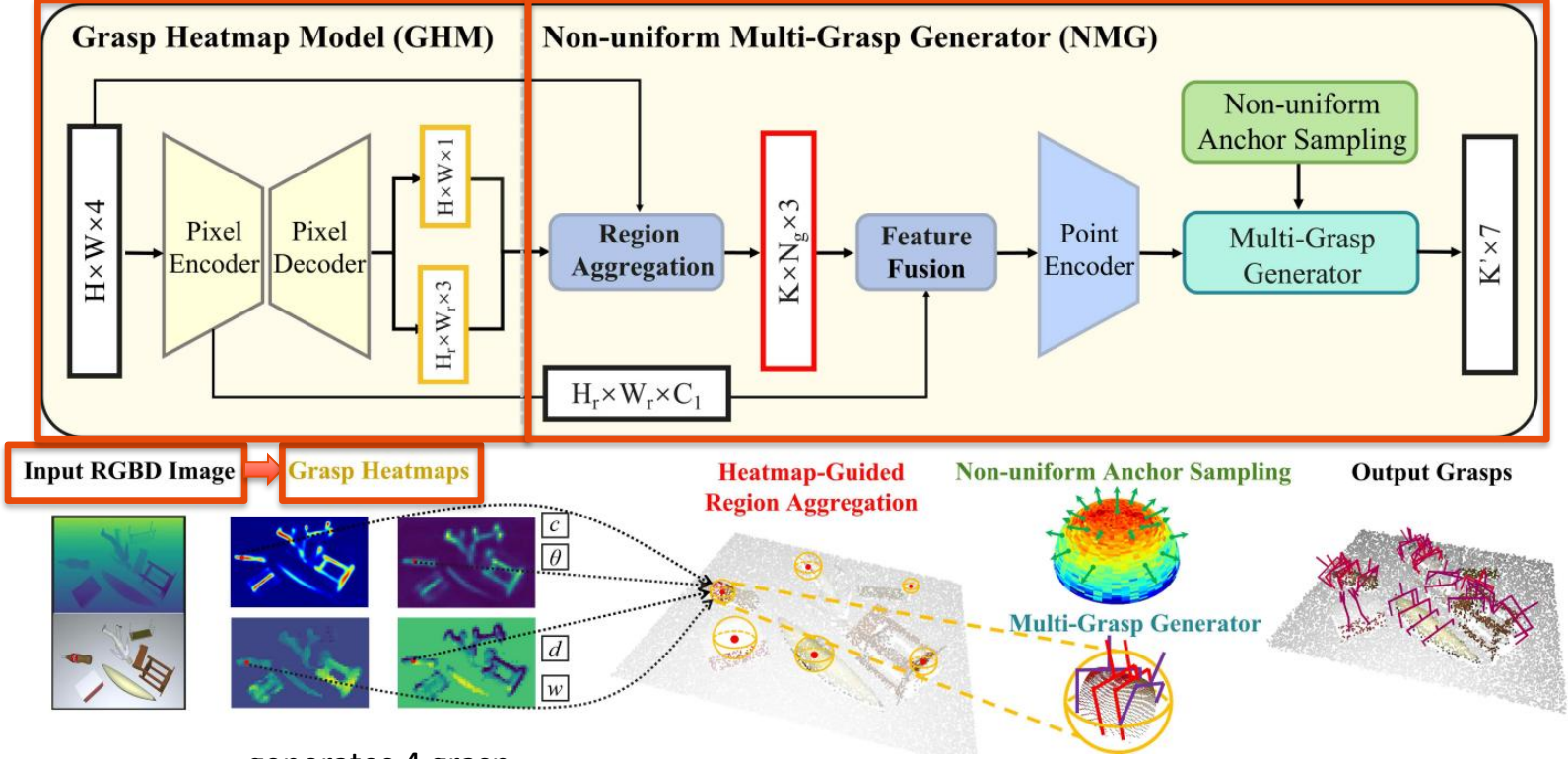


Encoder-decoder - extracts semantic features via a CNN



generates 4 grasp heatmaps as guidance

Encoder-decoder - extracts semantic features via a CNN

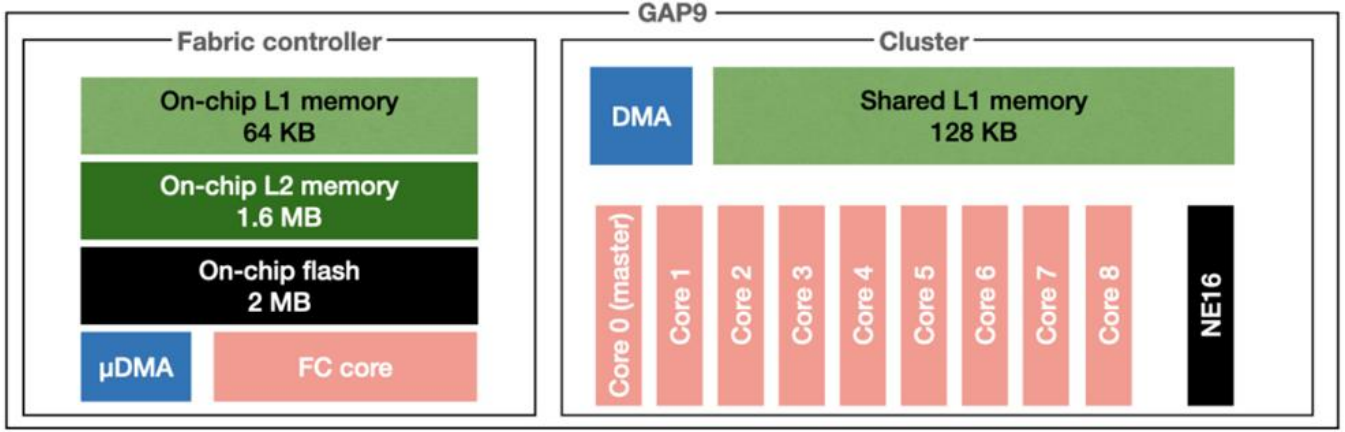


generates 4 grasp heatmaps as guidance

takes the heatmaps and the scene point cloud as input and aggregates multiple graspable local areas

Part B: The HGGD-MCU model

GAP9 deployment



Greenwaves Application Processor 9 - GAP9

- The SoC operates at an internal clock frequency of up to 370MHz, with DVFS
- Fabric Controller (FC) core orchestrates system-level operations
- Cluster of 9 **RISC-V** cores, identical specifications to FC. Execute independently, allowing concurrent processing of NN
- A dedicated accelerator core, NE16, further enhances NN inference by accelerating MACs

Quantization

- **Core idea:** map real-valued numbers to a finite set of discrete levels using a linear transformation
- Reduces the numerical precision of model parameters and activations, converting 32-bit FP into lower bit representations -> reduces model size
- **Added benefit:** integer operations consume less energy than their FP counterparts

- **Input size reduction:** 640×360 -> 320 × 160

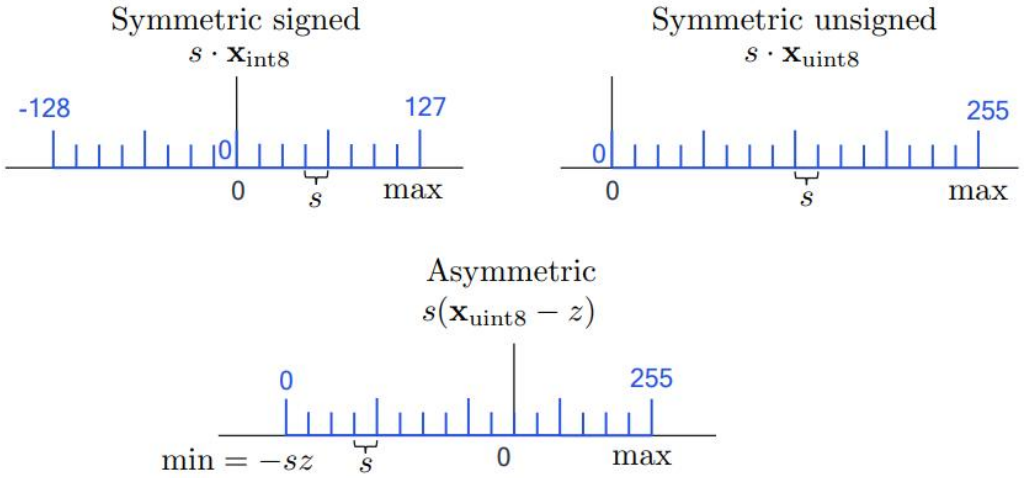
Integer	
Add	
8 bit	0.03pJ
32 bit	0.1pJ
Mult	
8 bit	0.2pJ
32 bit	3 pJ

FP	
FAdd	
16 bit	0.4pJ
32 bit	0.9pJ
FMult	
16 bit	1pJ
32 bit	4pJ

M. Horowitz, "computing's energy problem (and what we can do about it),"

Quantization

- **Core idea:** map real-valued numbers to a finite set of discrete levels using a linear transformation
- **Scale *s***, determines quantization granularity
- **Zero-point *z*** shifts the quantized range to best approximate 0 in the original domain



M. Nagel, M. Fournarakis, R. A. Amjad, Y. Bondarenko, M. Van Baalen, and T. Blankevoort, "A white paper on neural network quantization,"

Quantization

Chosen parameters:

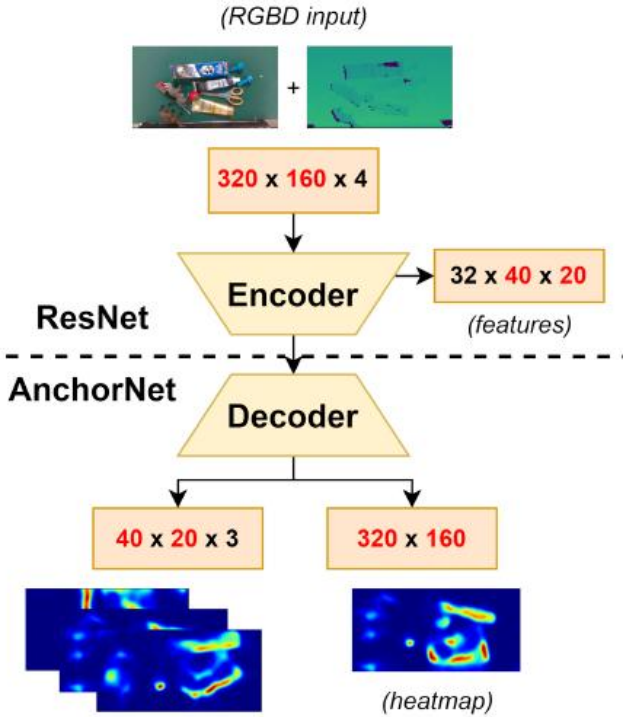
- *[Granularity]* - Per-tensor quantization, sharing *s* and *z* across the entire tensor

- *[Range estimation method]* - the strategy to define θ_{min} and θ_{max}. Used the min-max strategy, setting θ_{min} and θ_{max} to the minimum and maximum values of the input data to cover the whole dynamic range of the tensor

Pipelined execution

Chosen parameters:

- *Adapted architecture of HGGD*
- *Dotted lines represent partitioning points, where the original model is split into sub-models*



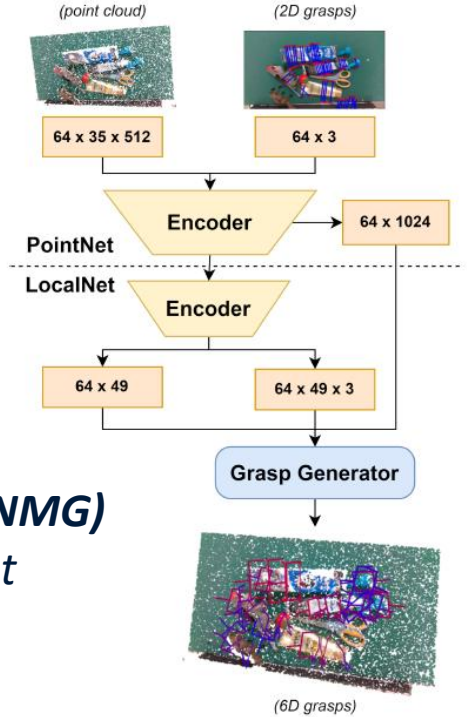
Grasp Heatmap Model (GHM)

Encoder-decoder -
extracts semantic features via a CNN

Pipelined execution

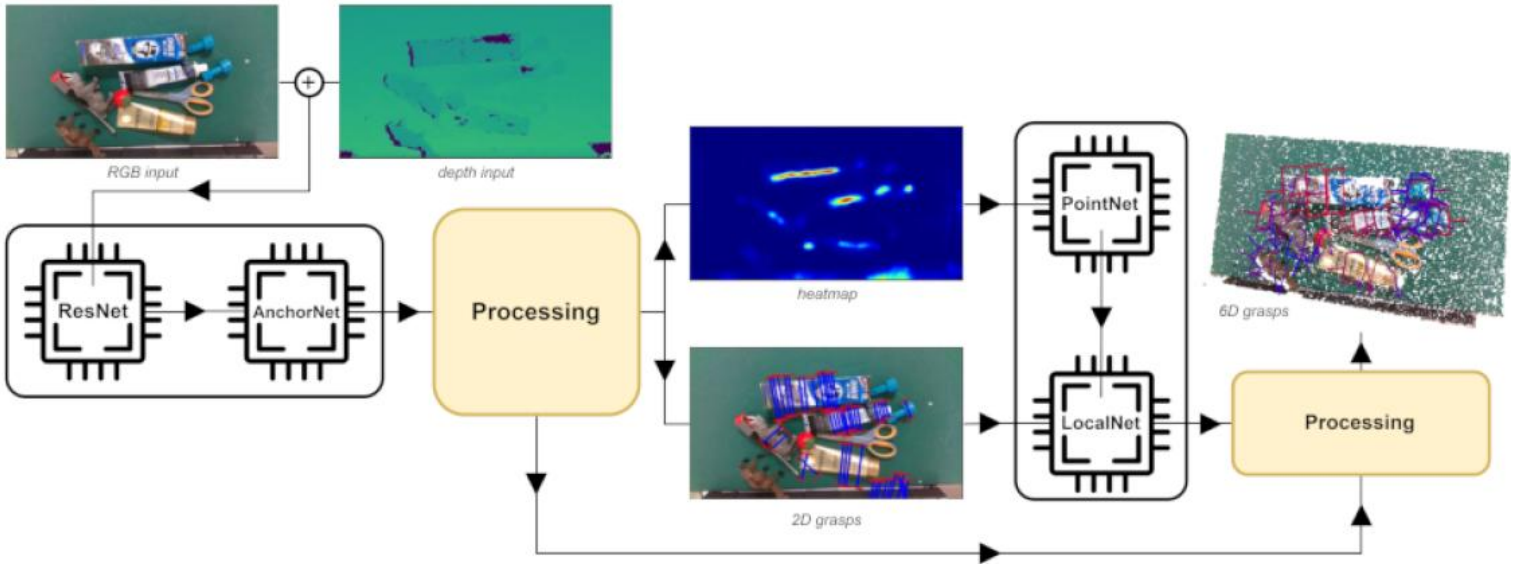
Chosen parameters:

- *Adapted architecture of HGGD*
- *Dotted lines represent partitioning points, where the original model is split into sub-models*

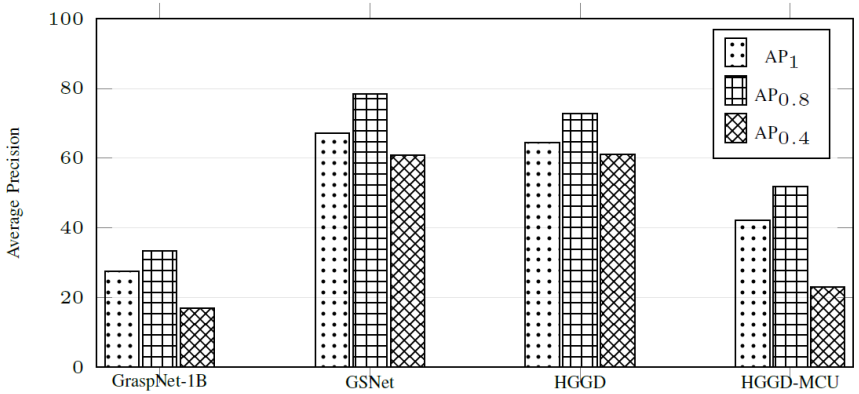


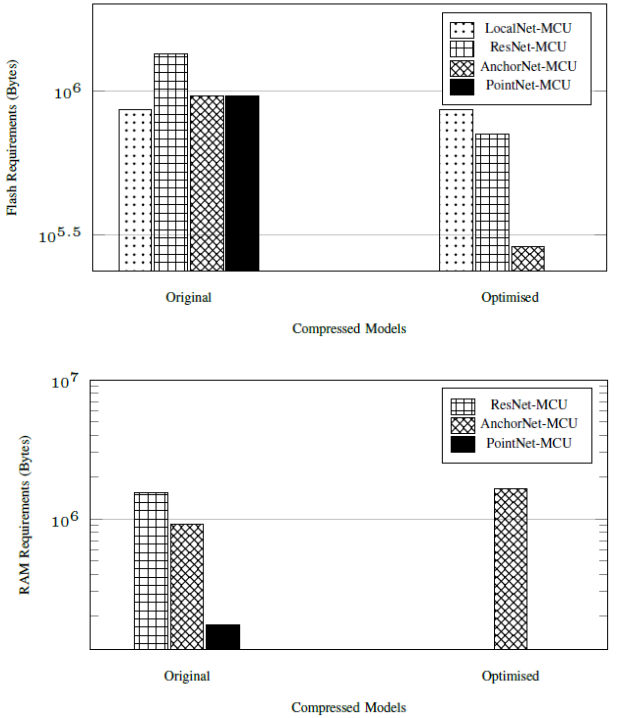
Non-uniform Multi-Grasp Generator (NMG)

Takes the heatmaps and the scene point cloud as input and aggregates multiple graspable local areas



Edge grasping perception pipeline





Average precision evaluation on the GraspNet-1Billion dataset

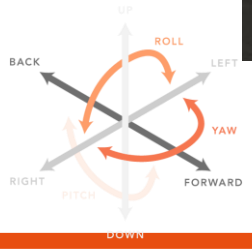
Hao-Shu Fang et. al., GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping

Conclusions

- Compressed HGGD using quantisation for execution in the GAP9 low-power edge device
- Split HGGD in 4 sub-models to enable pipelined execution

Future work

- Still work in progress!
- Investigate NAS and transfer learning
- Validate in actual robotic tasks using KUKA arms



Thank you!

SCAN ME

Dr. Charis Kouzinopoulos

charis.kouzinopoulos@maastrichtuniversity.nl
https://www.linkedin.com/in/kouzinopoulos/
https://kouzinopoulos.github.io/

Learn more about us at
www.maastrichtuniversity.nl/dacs