Deployment of CNN backbones in INCAME

CNN Model
- *.caffemodel
- *.prototxt

Data Quantization & Network analysis

Quantized Weights

Network Topology

Original Compilation

INCAME Compilation

Original Intruction Sequence

Locate interrupt position

Virtual-Intruction ISA generation

Interrupt Config file

Virtual-Instruction ISA (VI-ISA) Sequence (instruction.bin)

INCAME Runtime interrupt

DDR: spaces for Instruction

VI-ISA Instruction Sequences

Instruction Fetch

Instruction Arrangemetn Unit (IAU)

Interrupt Request

Robot Software with ROS

Original Intruction Sequence

Robot Software with ROS

- Node 1: Feature-Point Extraction (FE)
- Node 2: Place Recognistion (PR)
- Node 3: VO with Feature-Point
- Node ... : Exploration/ Navigation ...

Software Level

Task decomposition

CNN backbones
- VGG (FE)
- ResNet (PR)
- ...

Interruptible VI-ISA Instruction Sequences

Post-Processing
- SoftMax
- ...
- Norm
- Rank

CPU Tasks
- C++
- Python

INCAME Computation Level

Computaion Deployment

Interruptible CNN Accelerator

Special Logic

ScratchPad memory

Software Optimizing

ROS Runtime

Programmable Logic (PL, FPGA)
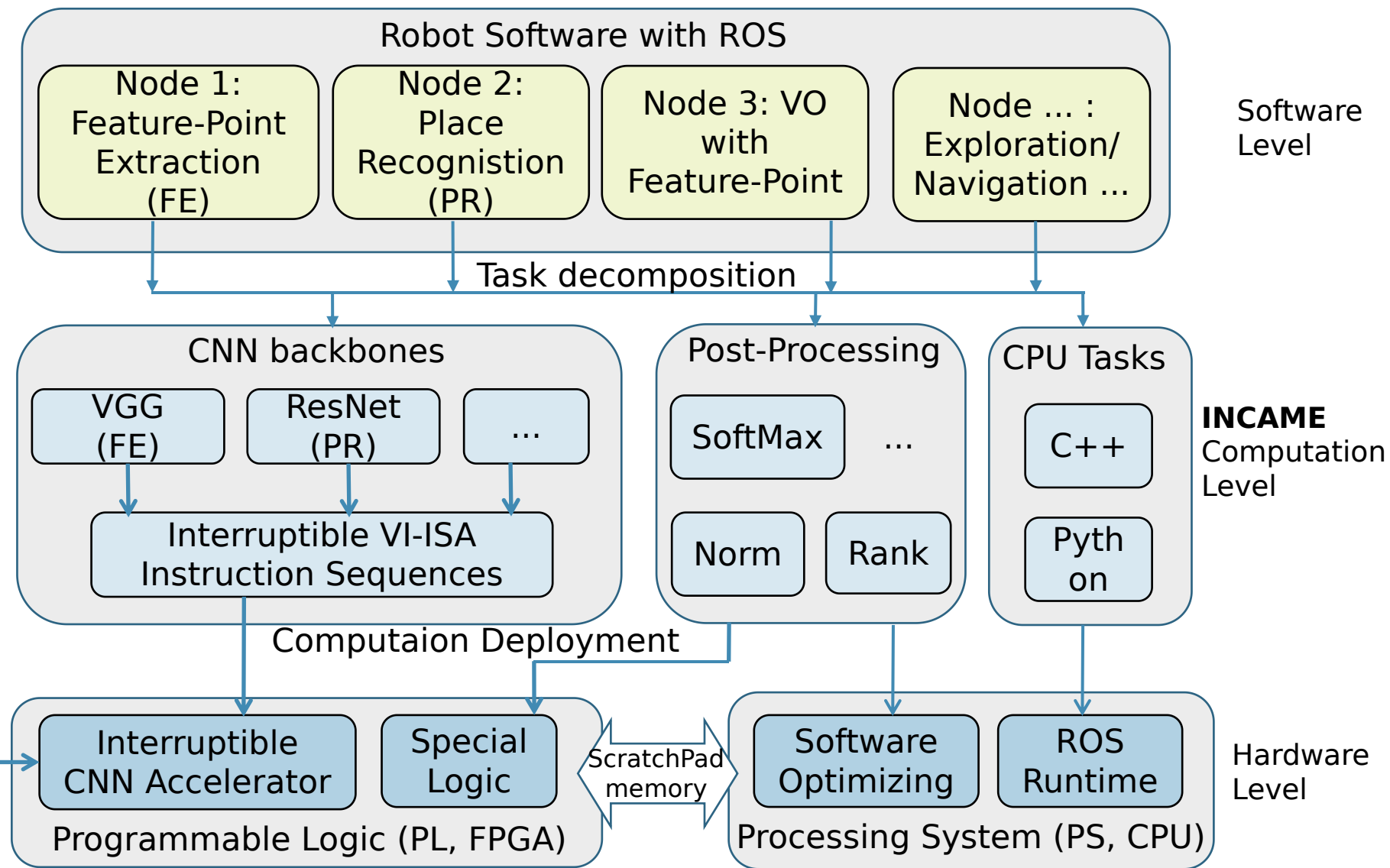
Processing System (PS, CPU)

Hardware Level

(a) The deployment of CNN backbones. At **compilation** step, INCAME locates the interrupt location, adds virtual instructions and genterates the virtual-instruction ISA (VI-ISA) sequence. At **runtime**, IAU translate the VI-ISA sequence to original sequence executed on the accelerator.

(b) The INCAME framework. At t**ask decomposition** step, the operations in different ROS nodes are seperated to CNN backbones, CNN Post-Processing, and Other CPU Tasks. At **computaion deployment**, the CNN backbone and Post-Processing are deployed with hardware modules and software optimizing.