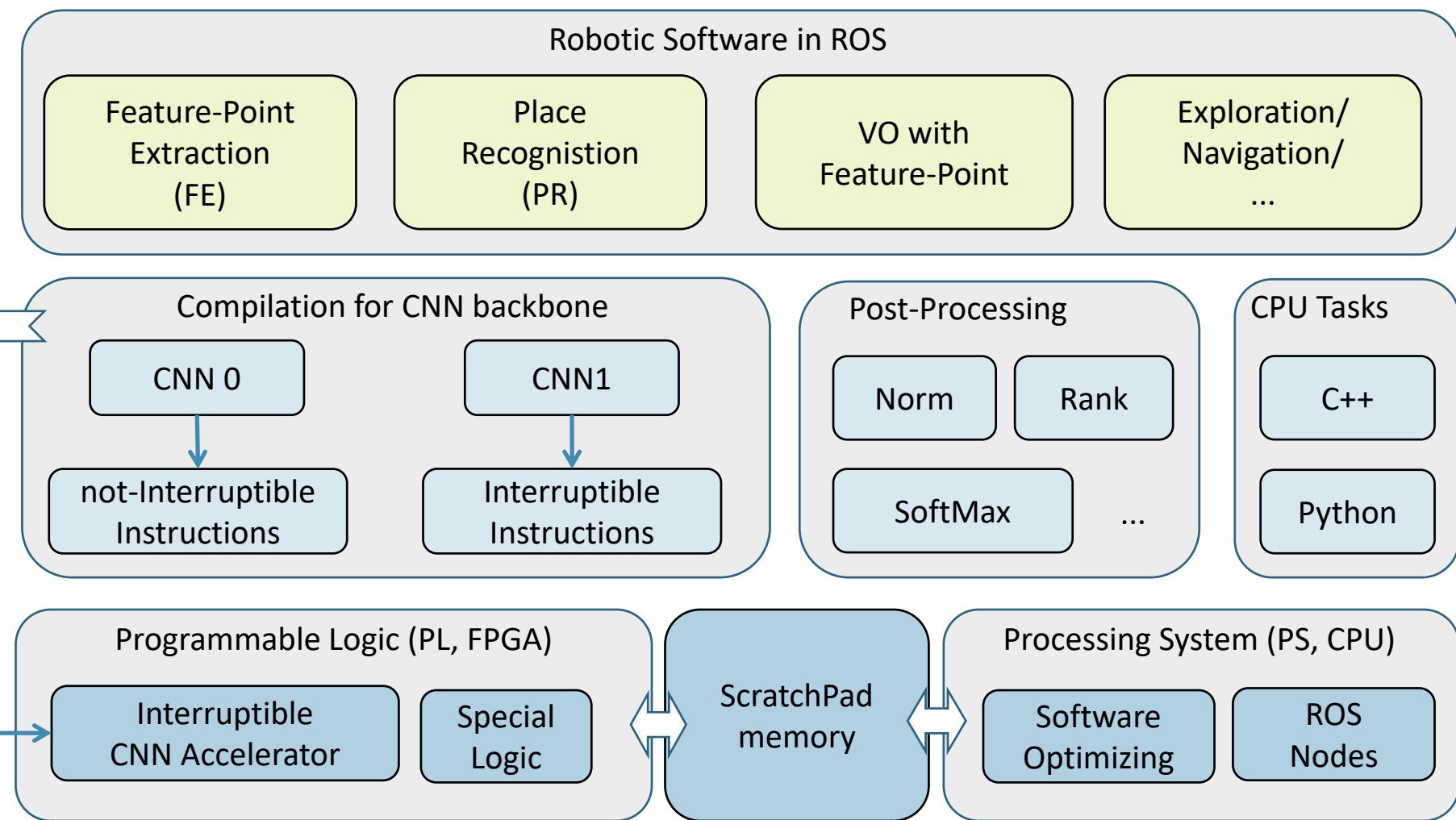


(a) The offline compilation with virtual-instruction ISA and the online interrupt between tasks. The normal instructions are generated from the CNN model, and then translated to virtual instruction ISA sequence. IAU translate the virtual instruction ISA sequence to normal sequence executed on the accelerator.



(b) The INCADE framework. The tasks (FE/PR/VO) are in different ROS nodes. The CNN backbone are translated to instructions, which enable runtime interrupt on the accelerator. The post-processing is accelerated by the special logic. Different modules on FPGA and CPU side share data through scratchpad memory.