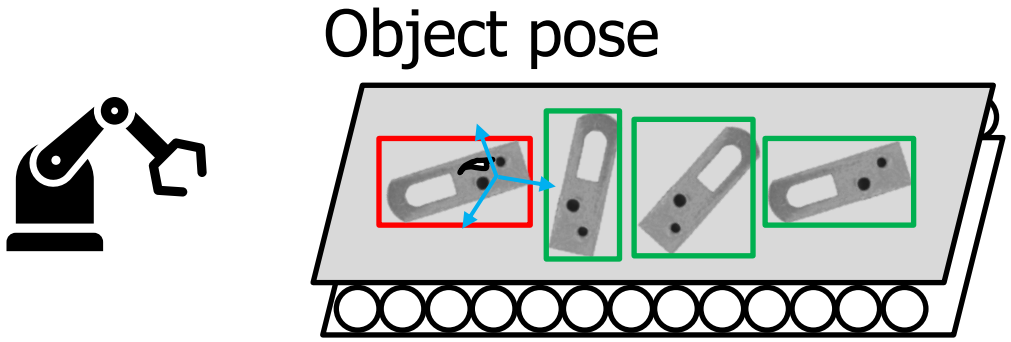
An ‘end to end’ 6DoF Object Pose Estimation Neural Network based on 1ms FPGA System

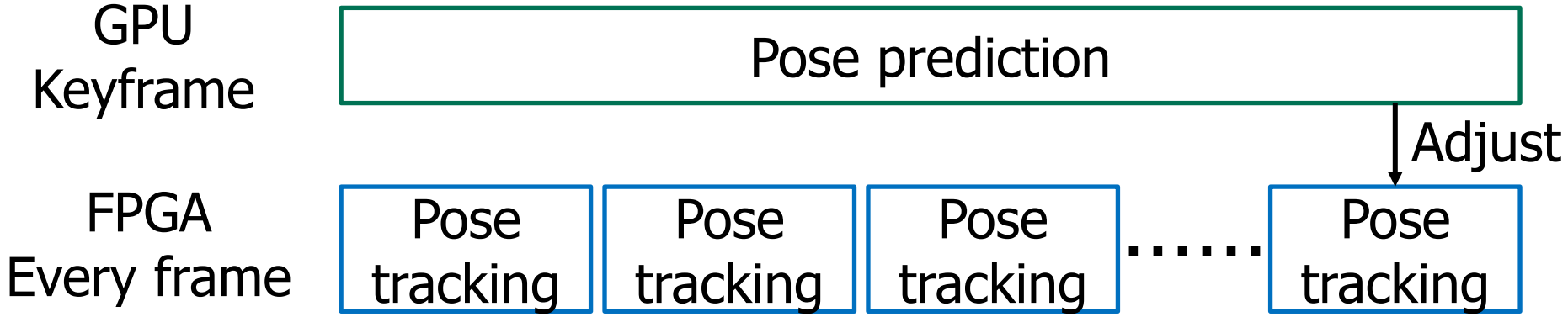
e2e method + RGB only + challenging scenarios = ???

2025/03/20

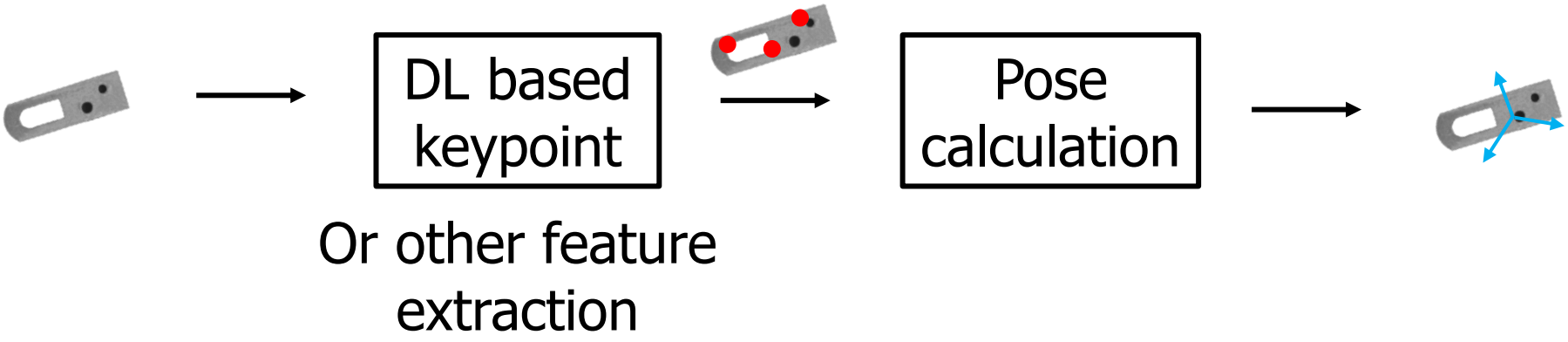
LU Weicheng

Inspiration





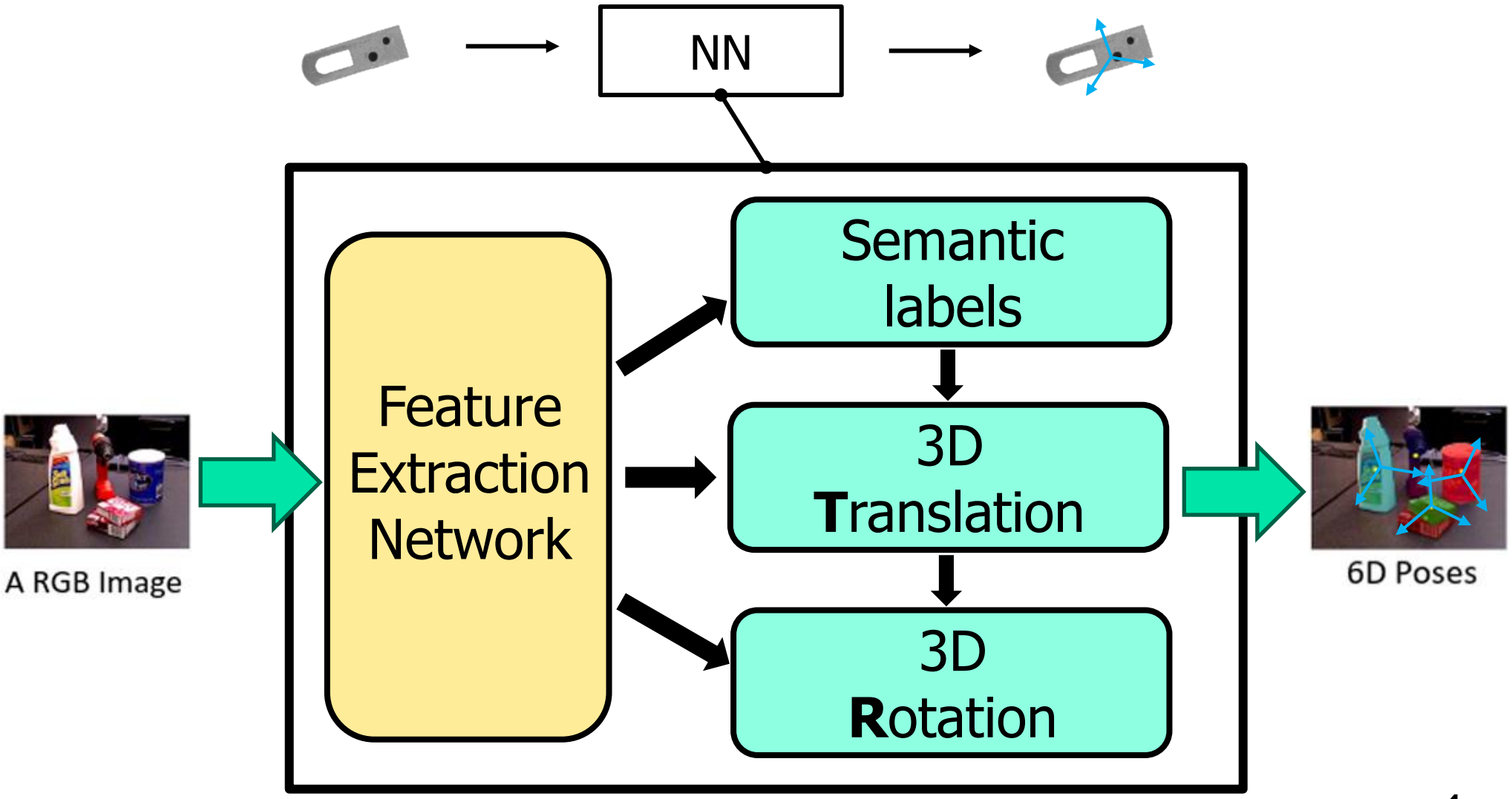
Keyword1: End-to-end Method



- Traditionally, after grasping keypoints by networks, we need to do one more mathematic calculation (such as PnP), in order to get pose estimation result.

Too slow **Costs too much resources**

Possible E2E-Network Structure



Keyword2: RGB based, No Depth

- Nowadays many studies use RGBD camera to realize and optimize 6D Object Pose Estimation, in order to reach higher accuracy.

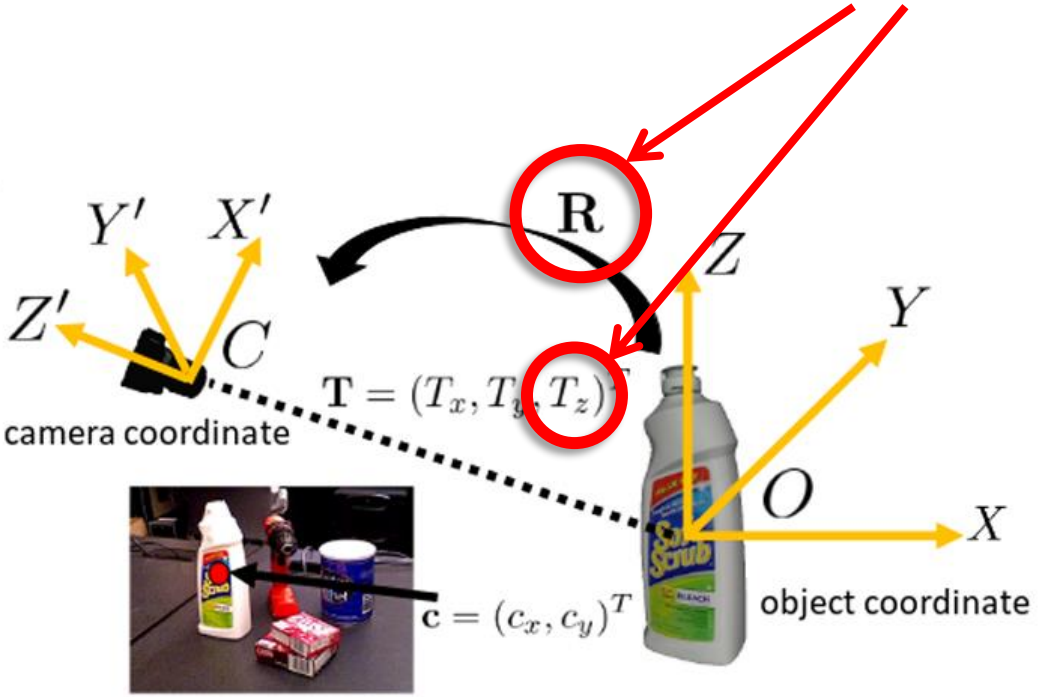
- However, considering 1ms FPGA system's super high-speed requirement and FPGA's on-chip memory, storing and processing objects' depth information costs too much.

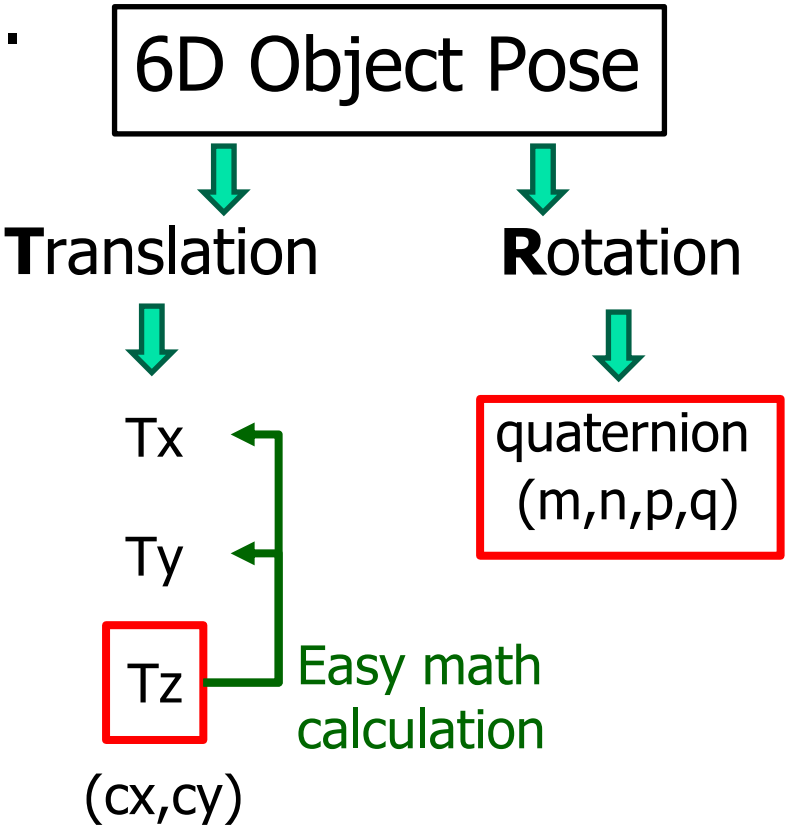
RGB-Only: otherwise >1ms

Just doing regression!

- Solution:

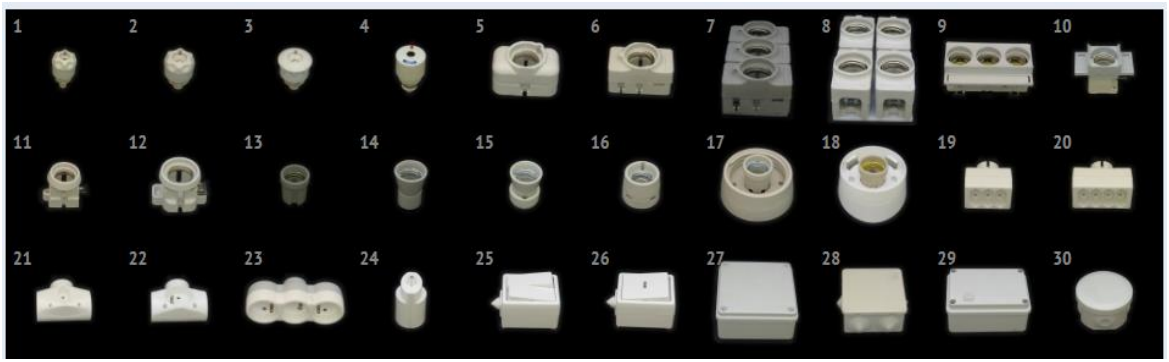
Directly use neural network to do both depth and rotation parameters' regression.





Keyword3: Challenging Scenarios

- Texture-less



- Symmetric

- Occlusion

- Transparent

Challenging but step by step

Realize under normal scenario (software side)

Start from the easiest

Realize under normal scenario in 1ms (hardware side)

Realize texture-less/symmetric (software side)

Realize texture-less/symmetric in 1ms (hardware side)

Realize transparent…

Deserves to have a try in the future

Thank You for Listening!