

Input: object point cloud,
end-effector point cloud
language instruction

Output: grasp configuration
(translation + rotation)



PointTransformerV3

⊕ element-wise add ❄️ frozen 🔥 active

Instruction
Grasp the handle of pan from
horizontal direction

