(a) Grasp Policy $\rightarrow (x_t, y_t, z_t)$ **Grasp Point Head** 73% 0.35 **Feature Pyramid Network Gripper Width Head** 3% **Vision Transformer Backbone** (+



