Table 1: Performance comparations between our STP and public VC-1 on **RLBench** simulation benchmarks. Both STP and VC-1 use **ViT-B/16**. Different from other simulation environments, we adopt a **multi-task setup** where all tasks share the same visual representation and policy module. We report the individual success rates for **20** tasks, as well as the average success rate, with each task comprising **25** evaluation episodes.

	STP	VC-1
insert_usb_in_computer	0	4
toilet_seat_down	100	96
sweep_to_dustpan	40	12
take_plate_off_colored_dish_rack	52	44
open_oven	32	0
stack_blocks	0	0
beat_the_buzz	12	12
open_wine_bottle	44	40
open_fridge	12	4
reach_target	92	68
basketball_in_hoop	96	100
put_rubbish_in_bin	92	88
meat_on_grill	92	32
take_frame_off_hanger	40	40
hang_frame_on_hanger	0	0
close_laptop_lid	52	60
take_usb_out_of_computer	100	88
change_clock	28	16
close_microwave	80	84
press_switch	76	76
Average	**52.0**	43.2

1