method	image forms	spatial tokens	non-spatial tokens	multimodal training	partial training
GLIGEN [19]	✗	✓	✗	✗	✓
T2I-Adapter [24]	✓	✗	✗	✗	✓
ControlNet [57]	✓	✗	✗	✗	▲
Uni-ControlNet [60]	✓	✗	✓	✓	▲
Composer [11]	✓	✗	✓	✓	✗
DIFFBLENDER	✓	✓	✓	✓	✓

"A cow grazes on grass in a pasture"

COW

Condition set

(a) SD ver.1.4 — text + "milk cow"

(b) GLIGEN — text + box

(c) ControlNet-S — text + sketch

(d) T2I-Adapter-D — text + depth

(e) PAIR-diffusion — text + style

(f) **DiffBlender** — all conditions