Figure 1: GLIGEN with different size of backbones all show a decline in aesthetic quality when conditional control increases.
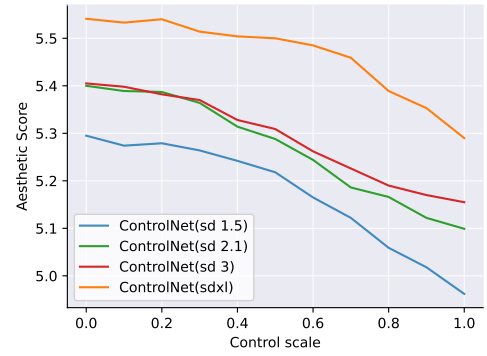

Figure 2: ControlNet with different size of backbones all show a decline in aesthetic quality when conditional control increases.
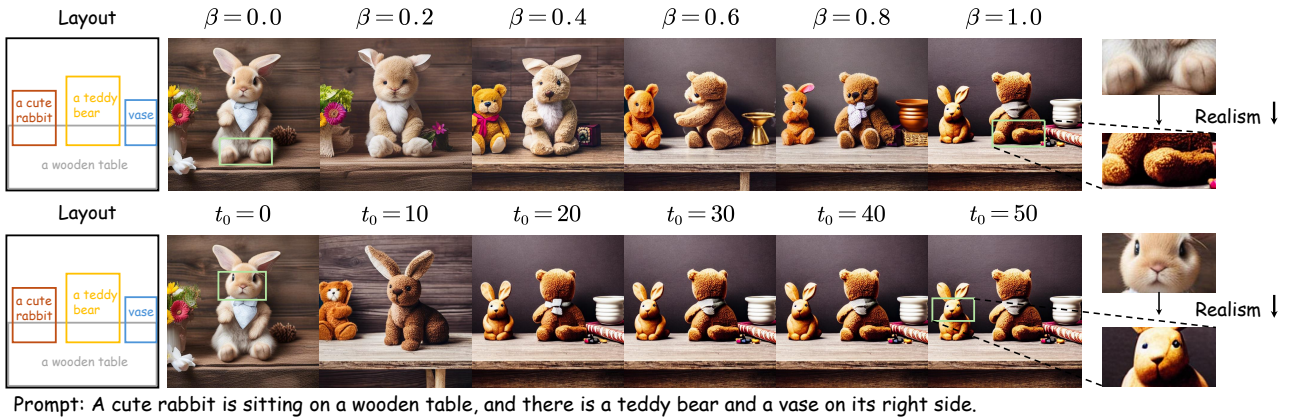
Model: GLIGEN

Prompt: A cute rabbit is sitting on a wooden table, and there is a teddy bear and a vase on its right side.

Figure 3: A more intuitive and clearer example to showcase our discoveries and motivation, using GLIGEN.

Model: InstanceDiffusion

Prompt: A yellow American robin, brown Maltipoo dog, a gray British Shorthair in a stream, alongside with trees and rocks.
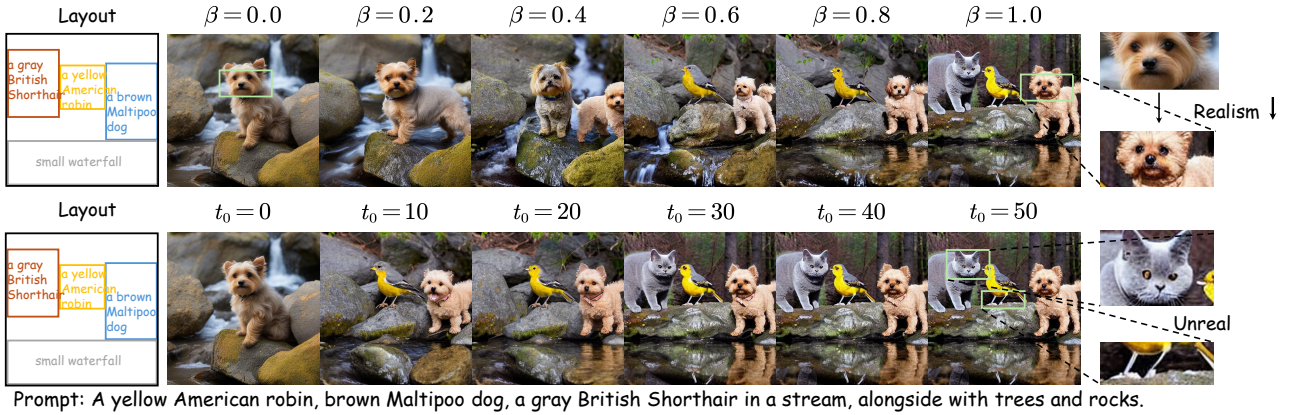
Figure 4: A more intuitive and clearer example to showcase our discoveries and motivation, using InstanceDiffusion.
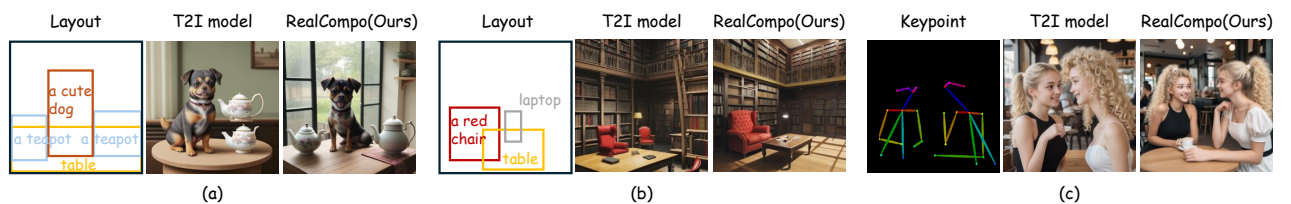

Figure 5: T2I models often generate unrealistic images due to unreasonable object positions. Our method improves image authenticity through conditional control while maintaining detail and aesthetic quality.