

Input

No prompt

Insufficient prompt  
(w/o mentioning "house")  
"high-quality and detailed masterpiece"

Conflicting prompt  
"delicious cake"

Perfect prompt  
"a house, high-quality,  
extremely detailed, 4K, HQ"

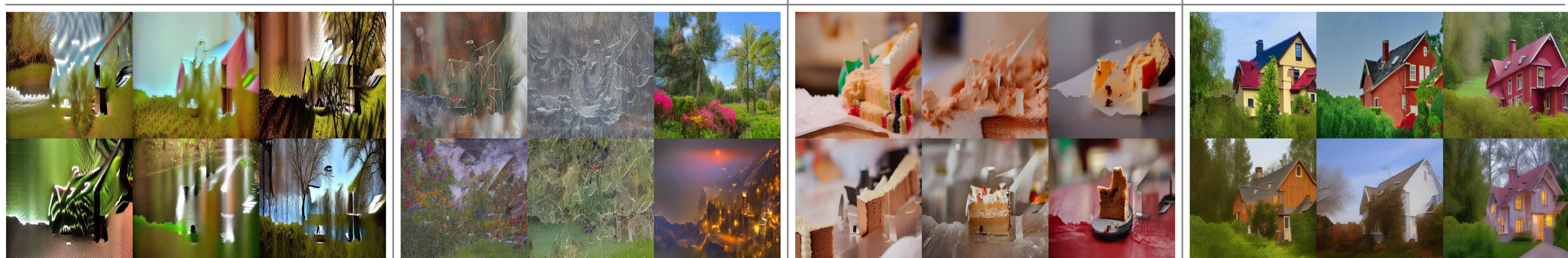
(a) Proposed



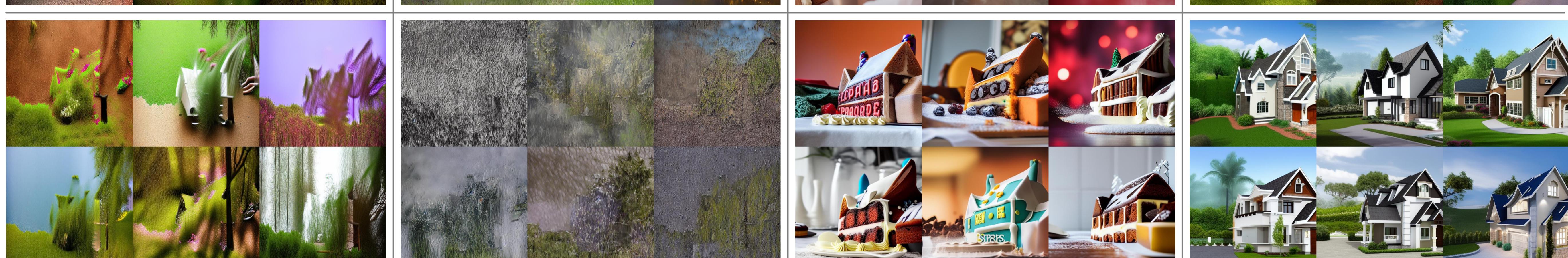
(b) w/o zero convolutions



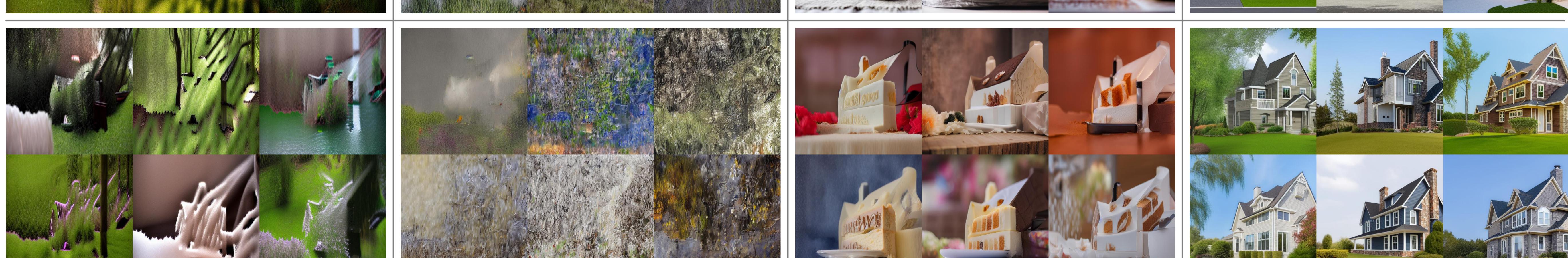
(c) w/o trainable copy,  
training lightweight  
layers from scratch  
(connecting encoder)



(d) w/o trainable copy,  
training lightweight  
layers from scratch  
(connecting decoder)



(e) w/o trainable copy,  
training lightweight  
layers from scratch  
(connecting decoder,  
using zero conv)



(f) directly train  
original model

