Toward Multimodal Image-to-Image Translation

Jun-Yan Zhu[1] Richard Zhang[1] Deepak Pathak[1] Trevor Darrell[1] Alexei A. Efros[1] Oliver Wang[2] Eli Shechtman[2]
[1]Berkeley Artificial Intelligence Research [2]Adobe Creative Intelligence Lab
https://github.com/junyanz/BicycleGAN

Problem Statement

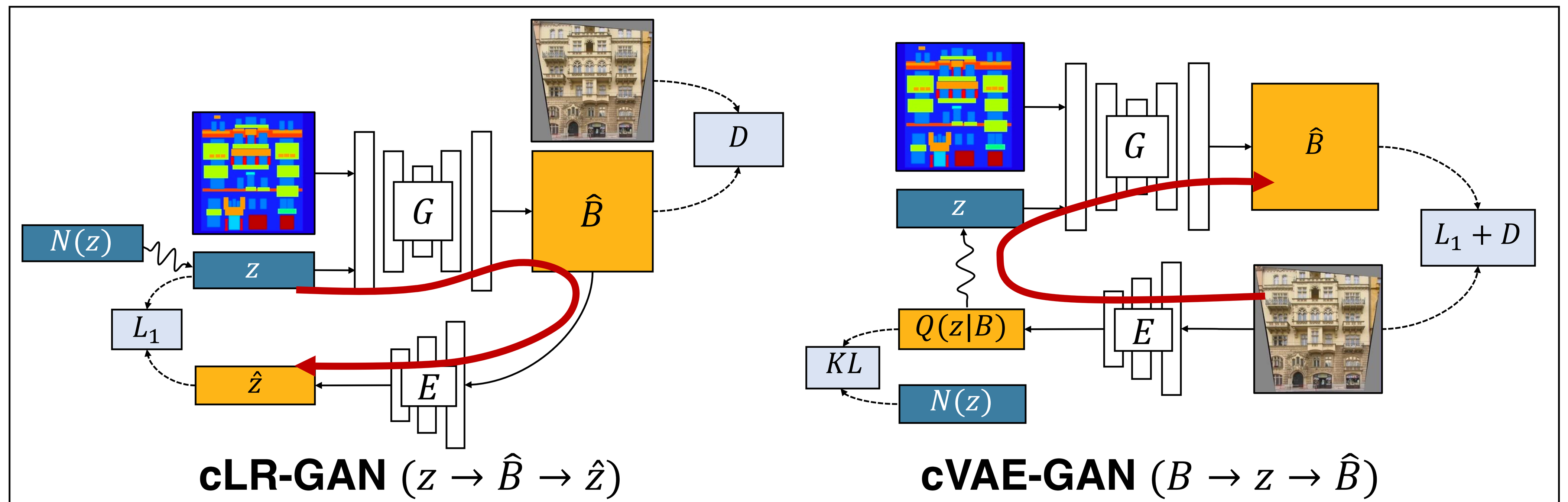
Develop general framework for one-to-many conditional image synthesis problems which produces ***diverse*** and ***realistic*** outputs

Our contributions

1) Detailed comparison of generative methods in conditional setting across a variety of datasets
2) Propose BicycleGAN, which encourages bijection between latent and output in multiple ways
3) Assess tradeoffs in latent space size and architectures

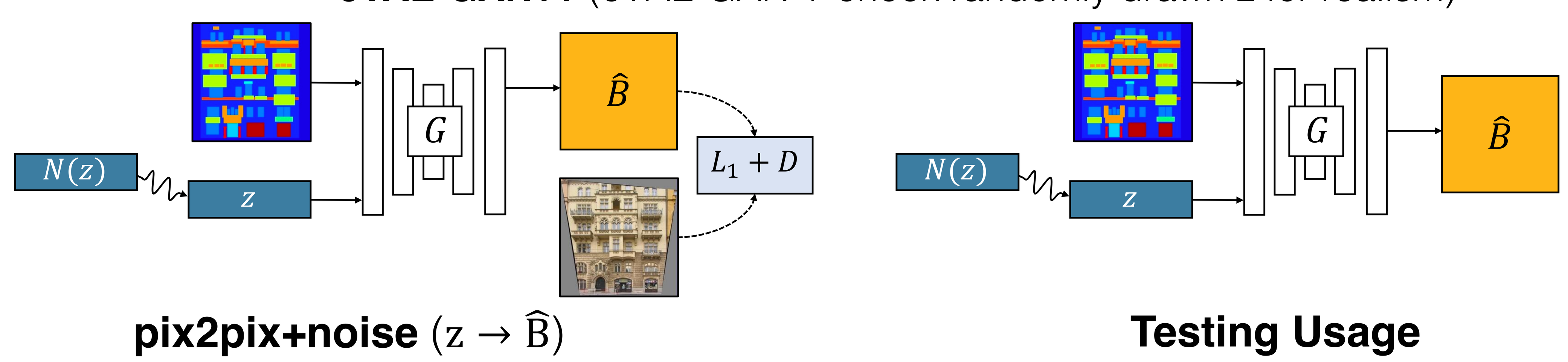
Example Qualitative Results

Methods



BicycleGAN (our full method)

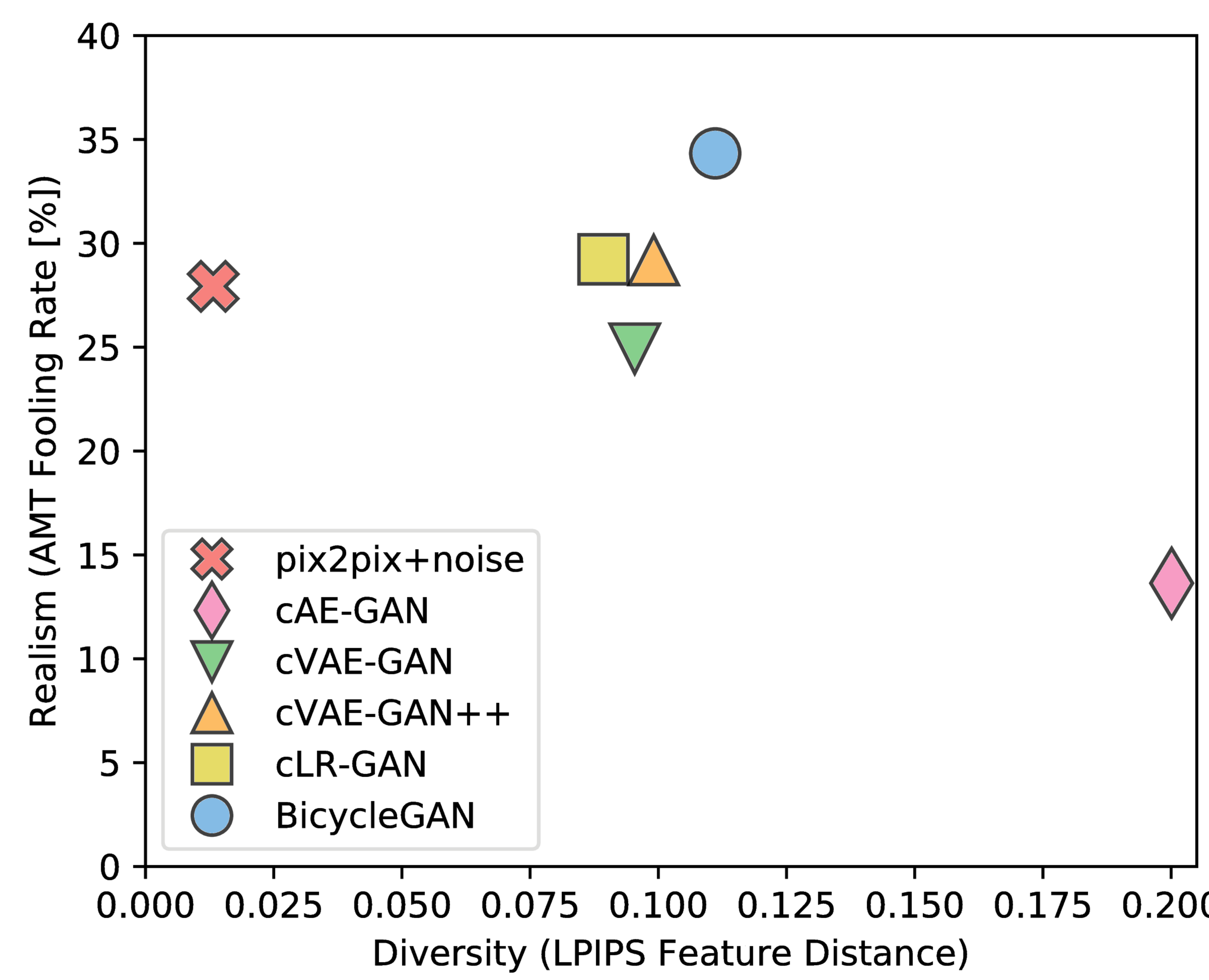
Other variants: **cAE-GAN** (cVAE-GAN without KL divergence on latent space)
cVAE-GAN++ (cVAE-GAN + check randomly drawn z for realism)



Realism vs Diversity

Assessment Metrics

- **Realism**: Real vs. Fake at Amazon Mechanical Turk
- **Diversity**: average feature distance between randomly drawn samples

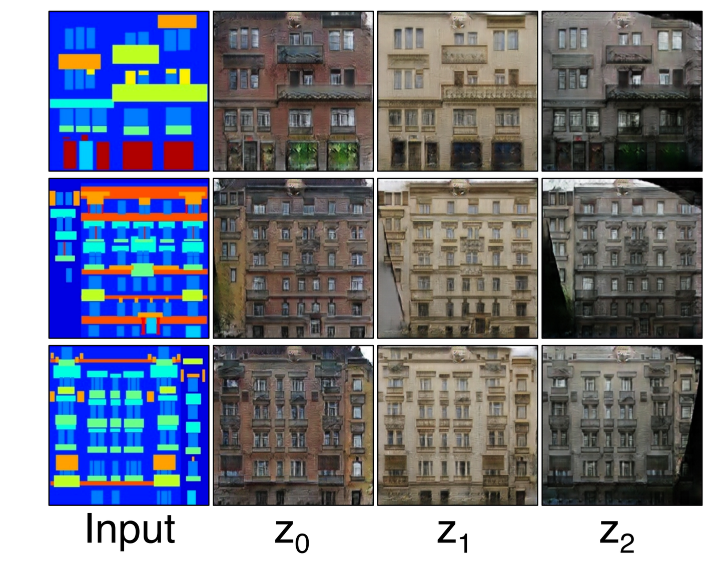


Conclusions

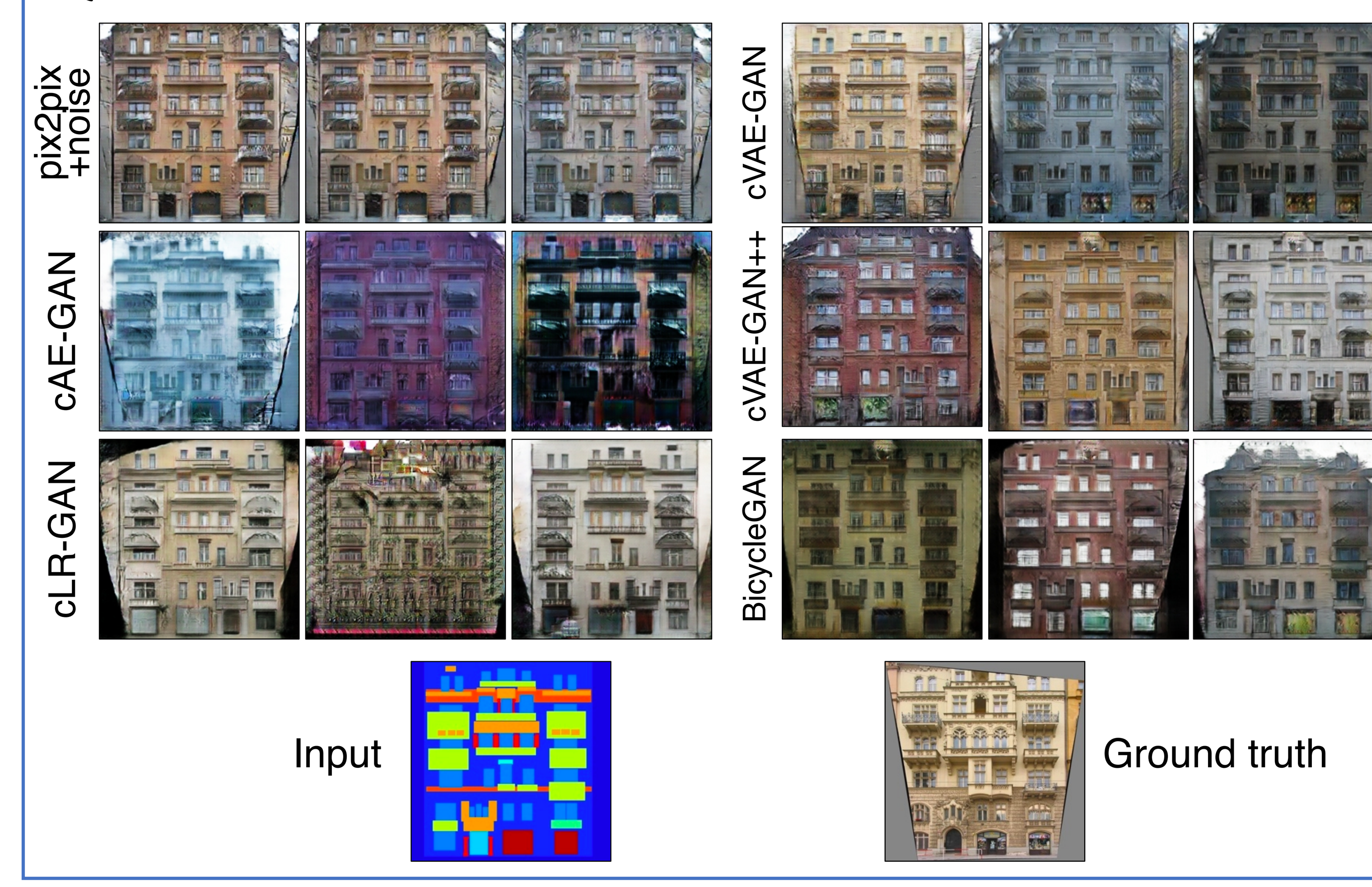
- pix2pix+noise baseline produces realistic results but little variation
- with no regularity in the latent space, cAE-GAN does not produce realistic samples
- checking randomly drawn samples helps (cVAE-GAN++ vs cVAE-GAN)
- combining cLR and cVAE-GAN into BicycleGAN helps realism and diversity

Latent Exploration

- Walking through the latent space
- Applying the same z across instances



Qualitative Comparison



Latent Space Size

larger |z| more expressive but difficult to densely fill

Encoder Arch, Method For Injecting z

Architecture

ResNet (E_ResNet) vs VGG-style Encoder (E_CNN)

Injecting z

input layer only vs every layer in the 1st half

Encoder	E_ResNet	E_ResNet	E_CNN	E_CNN
Injecting z	add_to_all	add_to_input	add_to_all	add_to_input
label→photo	0.292 ± 0.058	0.292 ± 0.054	0.326 ± 0.066	0.339 ± 0.069
map → satellite	0.268 ± 0.070	0.266 ± 0.068	0.287 ± 0.067	0.272 ± 0.069

Assess L1 reconstruction error