Image Neural Field Diffusion Models

Yinbo Chen[1], Oliver Wang[2], Richard Zhang[3], Eli Shechtman[3], Xiaolong Wang[1], Michael Gharbi[3]

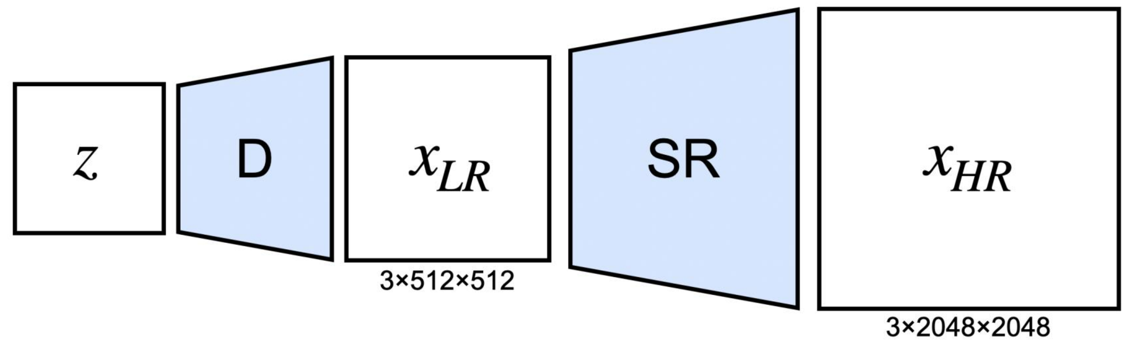
[1]UC San Diego [2]Google Research [3]Adobe Research

CVPR JUNE 17-21, 2024 SEATTLE, WA

Learning from Mixed and HR Datasets

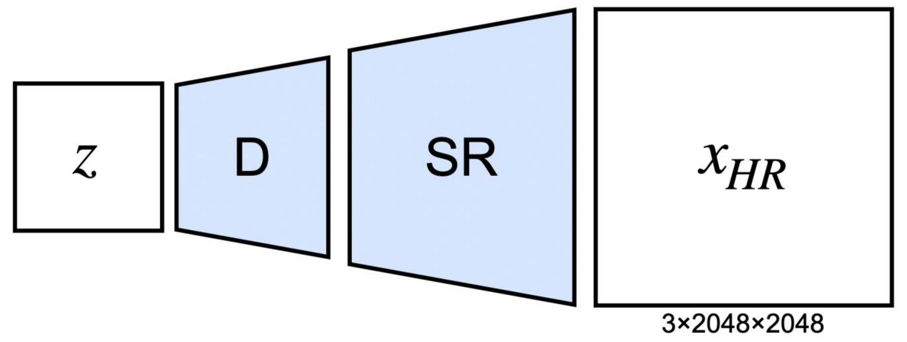
- Real-world images are at mixed and high resolutions
- Most prior methods downsample to fixed resolutions

1. Extra SR models



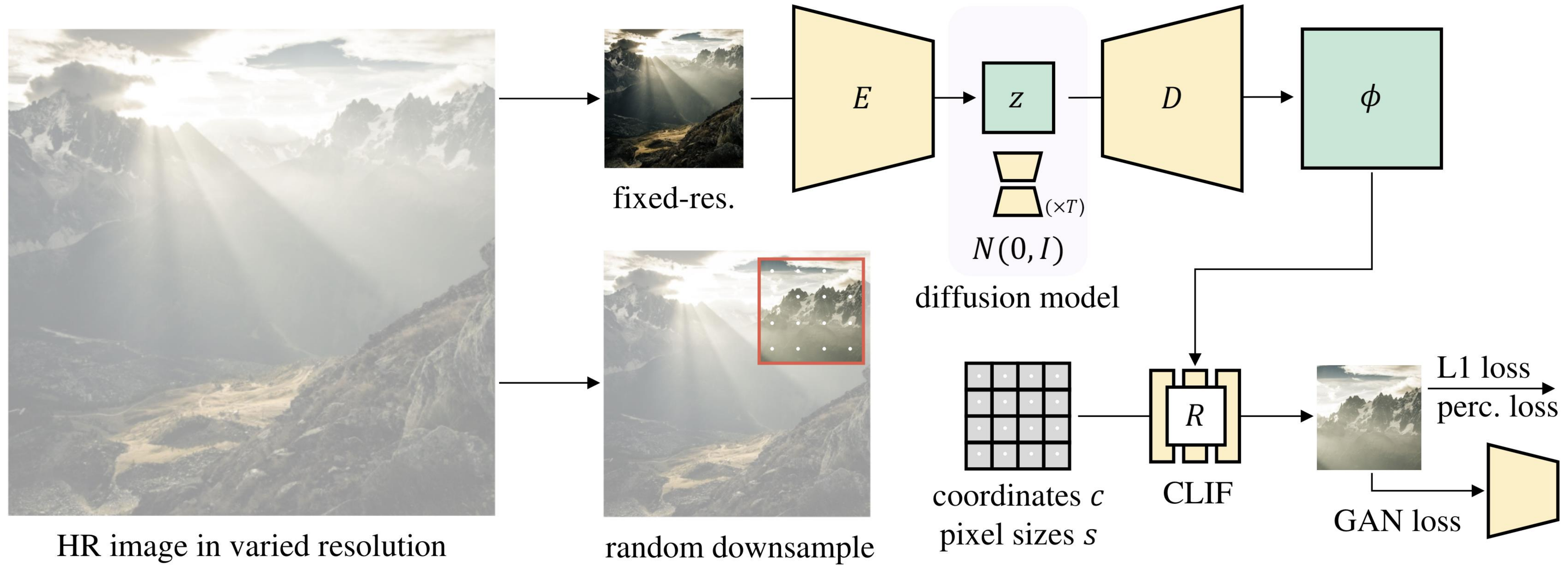
- LR is an info bottleneck, not robust to distribution shift

2. Remove bottleneck, direct to high resolution?



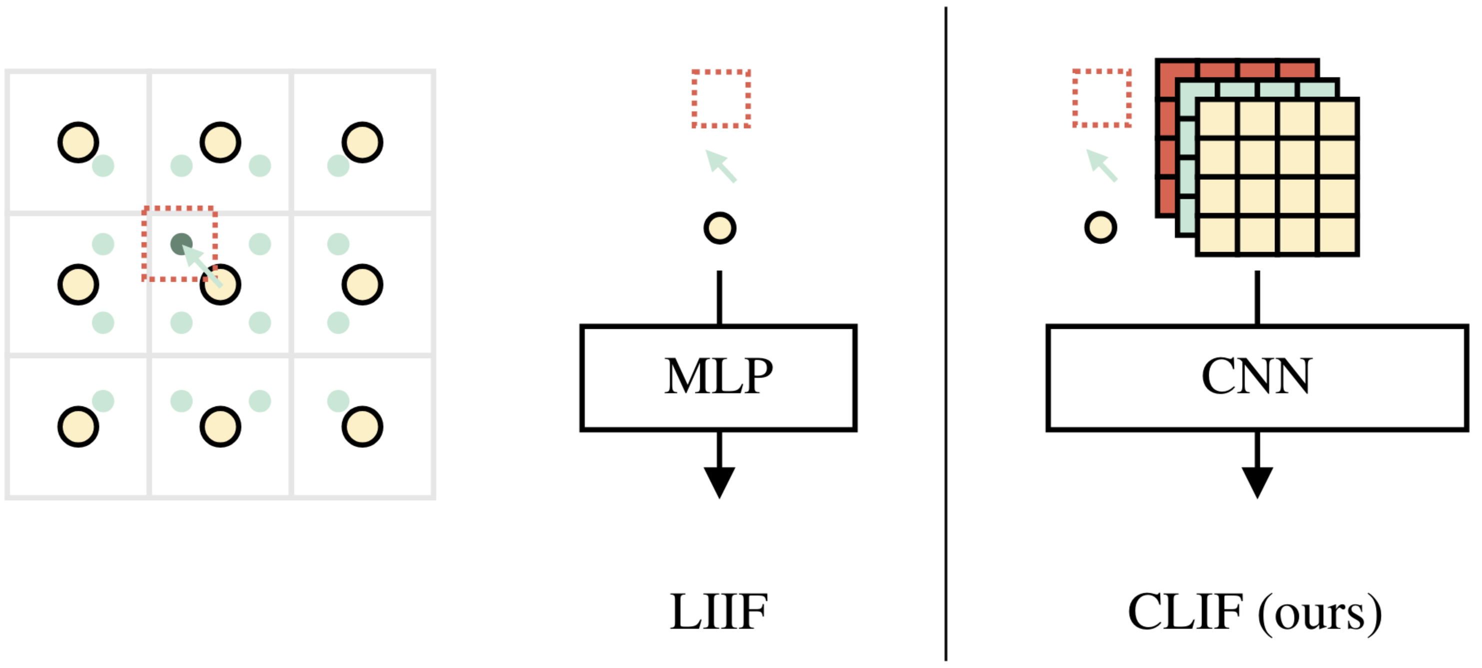
- Cannot learn from LR images (most web data)

Image Neural Field Diffusion Models (INFD)



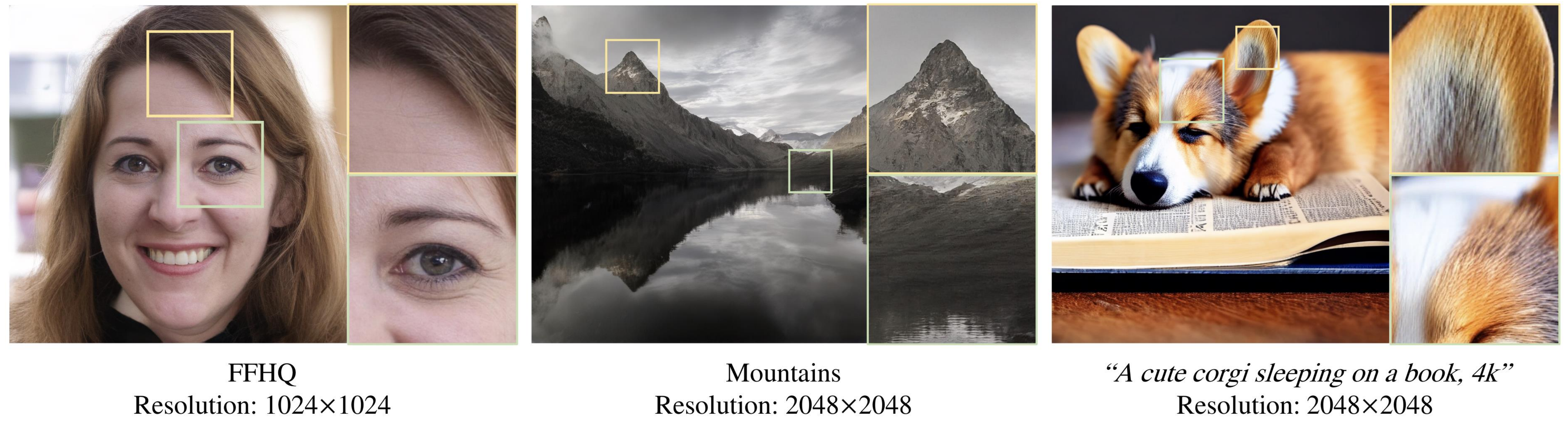
- A neural field autoencoder that maps image pixels to a photorealistic image neural field
- The representation is supervised by patches of images at arbitrary resolutions
- A latent diffusion model is learned to model the distribution of image neural fields

Convolutional Local Image Function (CLIF)



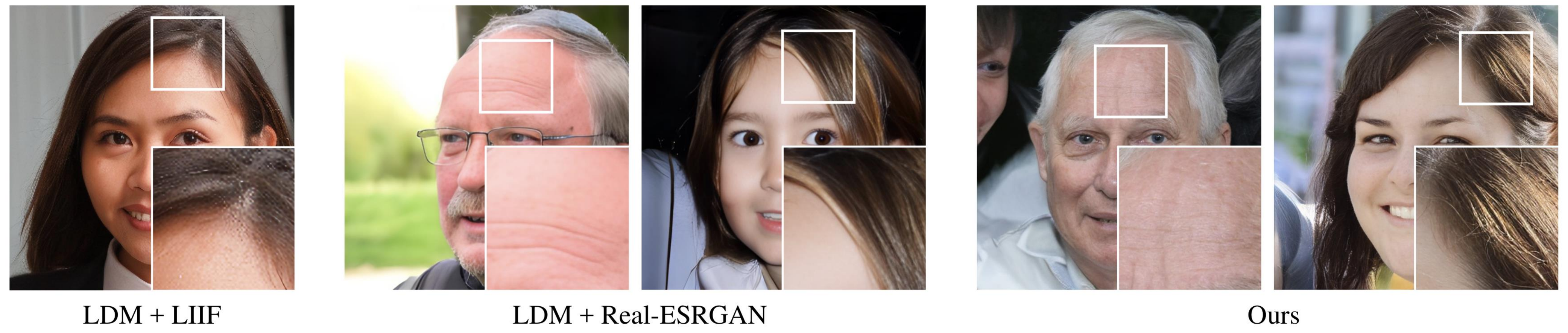
- Point-independent decoding constrains the design and is not sufficiently powerful
- All we need from image neural fields: patch rendering + scale consistency
- Use a ConvNet to decode the info map, scale consistency is observed after training

High-Resolution Generation in Different Domains



- Generate an image neural field and render it to a high resolution
- Diffusion can efficiently generate the compact latent (64×64 in examples)

Comparison to LR Diffusion + Extra SR Model



- Can learn from mixed-resolution training data and generate more details

Inverse Problems with Resolution-Agnostic Visual Prior

- Image neural field diffusion is a resolution-agnostic visual prior
- It can solve zero-shot inverse problems, with any-scale patch constraints
- Example: render given regions to 224x224 for CLIP similarity constraint