BuilDiff: 3D Building Shape Generation using Single-Image Conditional Point Cloud Diffusion Models

Yao Wei, George Vosselman, Michael Ying Yang
Scene Understanding Group, University of Twente, The Netherlands

AI3DCC @ ICCV23 PARIS

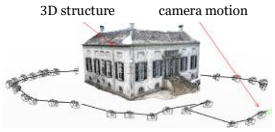
INTRODUCTION

- **Motivations**

Creating 3D building models typically requires tremendous manual efforts or expensive observations.


3D design

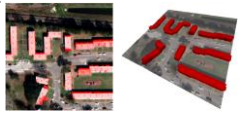

Airborne laser scanning



Multi-view reconstruction

Single image-to-3D building
- Generated 3D buildings exhibit LoD1 without roof structures.
- Images are limited to specific viewing angles.

Level-of-details (LoDs) defined by CityGML


3D building reconstruction from a single overhead image [CVPR'20]

- **Contributions**

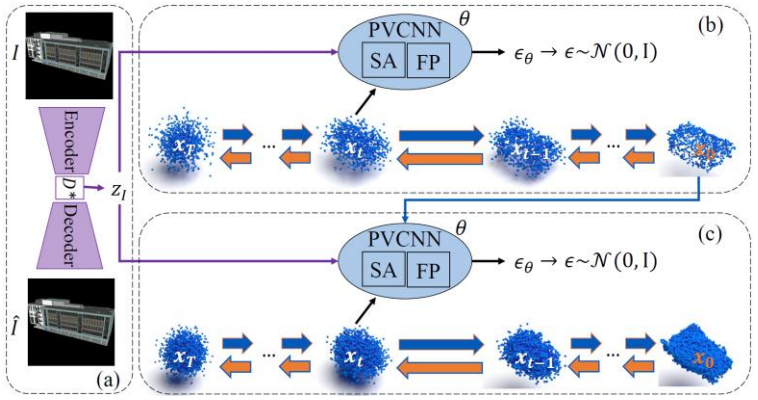
✓ A novel hierarchical framework is proposed to generate realistic 3D shapes of buildings with roof structures, i.e., at LoD2, given their single general-view images.

✓ Guided by an image auto-encoder, a base diffusion model coarsely identifies the overall structures of buildings, and an upsampler diffusion then derives higher resolution point clouds.

✓ A weighted building footprint-based regularization loss is introduced to constrain building structures and avoid ambiguous guidance during denoising process.

We are working on releasing our datasets and developing new benchmarks. For more about this project, please visit our homepage!

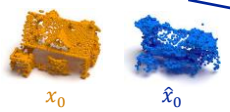
METHOD



- Image AE pre-training (AE)

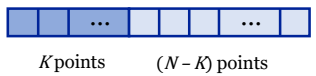
$$\mathcal{L}_{AE} = \mathcal{L}_{rec}(I, \hat{I}) + \mathcal{L}_{con}(z_I, z_I^a)$$

- Weighted building footprint-based regularization loss (WR)



$$\lambda(t) = \begin{cases} 1, & t = 1 \\ 0.75, & 1 < t \le T/4 \\ 0.50, & T/4 < t \le T/2 \\ 0.25, & T/2 < t \le 3T/4 \\ 0, & 3T/4 < t \le T \end{cases}$$

- Upsampler diffusion (UP)



Algorithm 1. Training of conditional diffusion

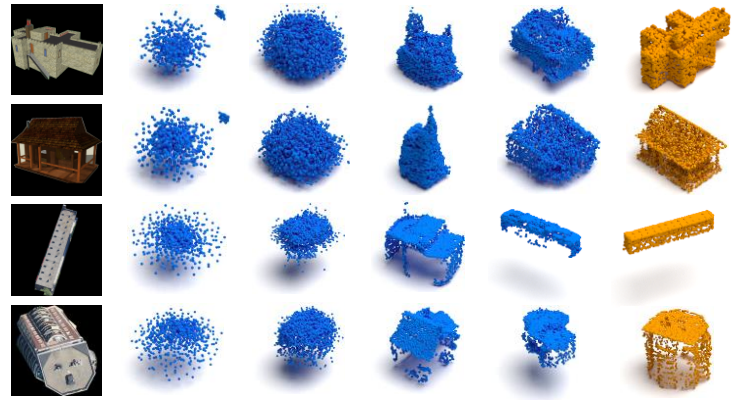
while not converge **do**
  sample $x_0 \sim q(x_0), \epsilon \sim \mathcal{N}(0, \mathbb{I})$
  sample $t \sim \mathcal{U}(\{1, \dots, T-1, T\})$
  $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$
  $\mathcal{L}_{eps} = \|\epsilon - \epsilon_\theta(x_t, t, z_I)\|^2$
  $\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t, z_I))$
  $\mathcal{L}_{reg} = \lambda(t)\Omega(proj(x_0), proj(\hat{x}_0))$
  $\mathcal{L}_\theta = \mathcal{L}_{eps} + \rho\mathcal{L}_{reg}$
  update model parameter θ with $\nabla_\theta \mathcal{L}_\theta$
end while

Algorithm 2. Sampling of conditional diffusion

sample $x_T \sim \mathcal{N}(0, \mathbb{I})$
for $t = T, T-1, \dots, 1$ **do**
  if $t > 1$ **then**
    sample $\boldsymbol{z} \sim \mathcal{N}(0, \mathbb{I})$
  else
    $\boldsymbol{z} = 0$
  end if
  $\epsilon_{guided} := (1+\gamma)\epsilon_\theta(x_t, t, z_I) - \gamma\epsilon_\theta(x_t, t, \emptyset)$
  $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{guided}(x_t, t, z_I)) + \sigma_t \boldsymbol{z}$
end for
return x_0

RESULT

Method	BuildingNet-SVI (Synthetic)			BuildingNL3D (Real)		
	CD↓	EMD↓	F1↑	CD↓	EMD↓	F1↑
MRTNet [ECCV'18]	6.11	49.07	6.89	2.84	44.06	5.18
FlowGAN [BMVC'22]	**2.00**	21.21	21.17	**2.33**	24.06	**22.26**
PVD [ICCV'21]	6.18	16.08	20.02	5.69	14.74	13.01
BuilDiff (Ours)	3.14	**10.84**	**21.41**	3.81	**10.43**	22.08



Ablation Study	CD↓	EMD↓	F1↑
Baseline@1024	11.109	24.027	5.180
AE@1024	6.406	16.770	8.531
AE_WR@1024	5.766	15.625	9.601
AE_WR@4096	4.046	13.373	15.227
AE_WR_UP@4096	3.810	10.427	22.081