FlexPart

3D Parts from Sparse Geometric Prompts

A Unified Framework for Image-to-3D Part Generation
via Variable Granularity

Qiwen Gu · Xin Zhou · Xiwu Chen
Zhongyang Zhu · Zhuo Zhang · Junqiao Zhao

Tongji University / Afari / HUST

Complex Parts from Simple Prompts

One image + a point per part, with optional 2D boxes



Points and boxes mark user prompts. Colors distinguish the generated parts.

Dense masks are optional

One Model, Multiple Prompts

Prompt	Spatial Information	Usage
Point	Part location	Base input
2D Box	Coarse extent	Optional cue
Mask	Visible boundary	Optional cue

Shared Learning

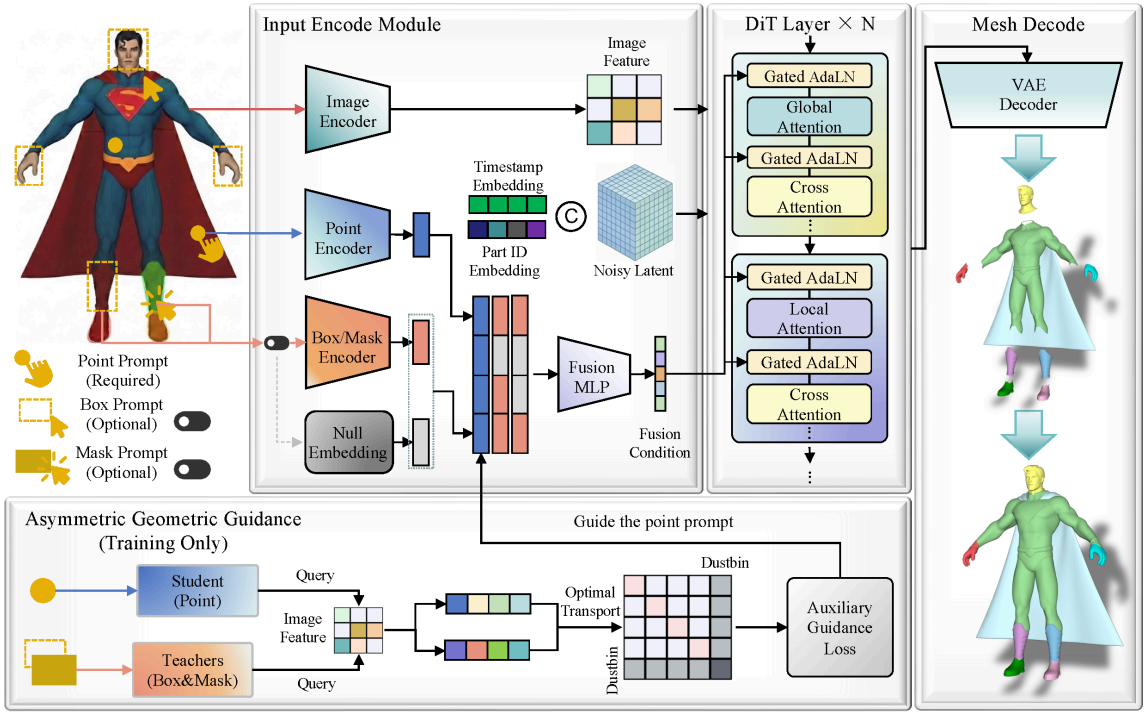
Flexible inputs

Complementary cues

Shared geometry

Can dense prompts improve part structure under sparse inputs?

FlexPart Architecture



Prompt encoding and fusion

Gated AdaLN conditioning

Global / local attention

Guidance branch: training only

Unified Geometric Conditioning

Prompt	Lightweight Encoder	Spatial Information
Point	Fourier features + MLP	Continuous 2D position
2D Box	Corner encoding + corner-type embeddings	Part extent
Mask	CNN + global average pooling	Visible contour

Learnable null embeddings handle missing prompts.
A fusion MLP produces a fixed-size condition.

Gated AdaLN predicts scale, shift and gate at each layer.

Asymmetric Geometric Guidance

Role	Condition	Learning Objective
Teacher	Box / Mask	Provide stronger spatial cues
Student	Point	Resolve point ambiguity
Matching	Partial OT	Match across conditions

Shared visual bridge
Image context links prompt features

Asymmetric updates
Stop-gradient on teacher prompts

A dustbin absorbs unmatched mass when conditions are missing.

Training only: no teachers or alignment at inference

Part Generation Quality

Point / Box Part CD: 0.3334 / 0.3042 on PartVerseXL

Method	Input	PV CD ↓	PV IoU_B ↑	Tiny CD ↓	Tiny IoU_B ↑
TRELLIS + SAM3D	MV Mask	.3471	.2173	.3916	.2031
TripoSG + SAM3D	MV Mask	.3523	.1896	.4235	.2078
PartCrafter	Count	.5235	.1257	.5193	.1455
PartPacker	Image	.3933	.2210	.3932	.2327
OmniPart	Mask	.3283	.2831	.3681	.2649
FlexPart	**Point**	**.3334**	**.2506**	**.3802**	**.2456**
FlexPart	**Box**	**.3042**	**.2810**	**.3675**	**.2693**
FlexPart	**Mask**	**.2938**	**.2840**	**.3560**	**.2778**

PV = PartVerseXL. Tiny = PartObjaverse-Tiny. Inputs differ across systems.
Box / Mask also use points. Scores cover matched visible parts only.

Joint Training and Guidance

Joint Training Reduces Part CD

Prompt	Single	Joint	Drop
Point	**.3550**	**.3334**	**6.1%**
Box	**.3155**	**.3042**	**3.6%**
Mask	**.3224**	**.2938**	**8.9%**

Guidance for Point Prompts

Point Input	CD ↓	IoU_B ↑
No guidance	.3406	.2323
With guidance	**.3334**	**.2506**

7.9% Relative Gain in IoU_B

Shared training improves geometry under sparse prompts.

PartVerseXL: single vs. joint training (left), guidance ablation in the joint model (right).

Complex Part Generation Examples

Clothing, armor and accessories with manual point / box prompts



Shoulder armor, cloak and axe



Headgear, clothing layers, sword

Qualitative examples from the paper. Colors distinguish parts. Occlusion remains a limitation.

TIEV 途灵
Tongji Intelligence, Embodiment & Vision

Unified Part Generation with Sparse Prompts

One image + point prompts, with optional 2D boxes

Unified conditioning and geometric guidance
improve generation under sparse prompts

Limitations: single-view occlusion and depth ambiguity

github.com/Gucci233/FlexPart