separation

PRISM

Position-guided Region-based Image Separation into Multi-layers

ICPR 2026 Oral Presentation

Shuai Pan*, Xu Zhang*, Jiayin Chen, Wei Zhang

Advanced Institute of Information Technology, Peking University
* Equal Contribution

Roadmap

ORAL STRUCTURE

01

Problem & Motivation

Why editable layers require more than visible segmentation

02

PRISM Method

Detect → condition → generate → decode into RGBA layers

03

Experiments

Quantitative, qualitative, and ablation evidence

04

Conclusion

Complete, coherent, and editing-ready transparent layers

Motivation: From Flat Images to Editable Layers

PROBLEM

Creative editing needs independent, complete object layers — not only visible pixels.

User need
Move, recolor, replace, or recombine each scene element independently.

Core obstacle
Occlusions hide object parts; visible masks cannot recover missing structure.

Target output
Single RGB image → background + complete RGBA element layers.

Flat RGB image → **visible region only** → **complete editable layers**

- Layer decomposition must infer occluded content while matching visible texture.
- Boundaries must be smooth enough for direct downstream editing.

Task Definition: Multi-layer RGBA Decomposition

FIG. 1

PRISM turns a single composited image into a background layer and independent transparent element layers.

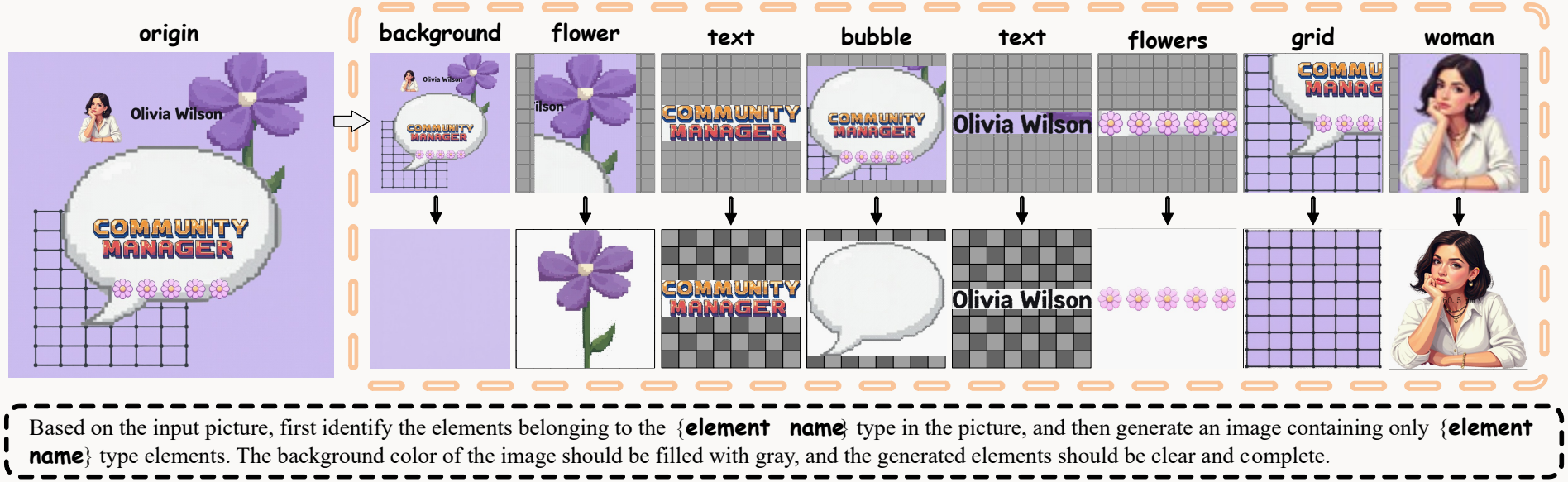


Fig. 1: element-level transparent layer decomposition with complete alpha channels

Key implication

Each output layer preserves object identity and structural fidelity, enabling independent editing.

Comparison with Previous Works

RELATED WORK

The key gap is not “finding pixels”; it is reconstructing complete editable layers under occlusion.

Work family	What it provides	Main limitation	PRISM shift
Segmentation / SAM	Visible object masks	No hidden structure; binary edges	Use detection as spatial guidance
Image matting	Alpha refinement	Needs trimaps/strong priors; incomplete objects	Decode alpha after reconstruction
Text-to-layer generation	Transparent layer synthesis	Prompt ambiguity; weak image fidelity	Condition on observed regions
LayerD / LayerDiffuse	**Graphic layer decomposition**	**Artifacts, noisy boundaries, limited foreground disentanglement**	**Detect–generate–decode per element**

Takeaway
PRISM reframes layer extraction as region-conditioned reconstruction: recover what should be there, not only what is visible.

PRISM Overview: Detect → Generate → Decode

METHOD

A detection-guided pipeline injects instance-level visual conditions into a frozen FLUX backbone.

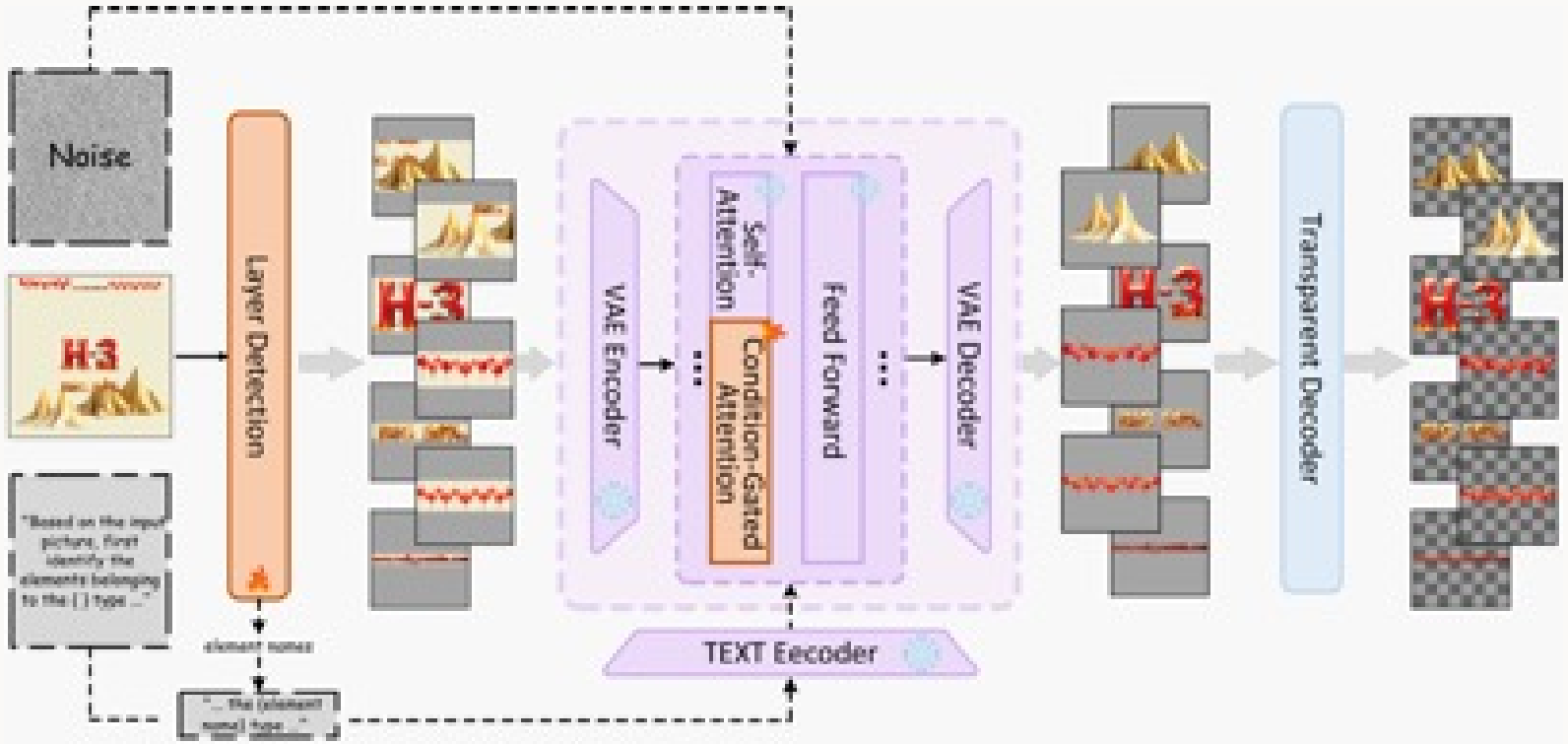


Fig. 2: overall architecture; only detection module and LoRA parameters are trained

Layer Detection

Condition-Gated LoRA

Transparent Decoder

Stage 1: Position-guided Layer Detection

METHOD 1/3

Detection supplies a localized visual condition for each element before generation.

Input
RGB image I

→

Detectron2
multi-scale proposals
+ RoI Align

→

Object layers
L = {L1, L2, ..., LK}

→

Prompted condition
element name +
cropped region

Why it matters
Spatial localization narrows the generation problem: each object is reconstructed independently while sharing the same diffusion timeline.

Resulting benefit
The model can complete occluded structures without forcing all elements into one global foreground mask.

Stage 2: Condition-Gated LoRA Attention

Low-rank adaptation is applied only to conditional image tokens, preserving FLUX priors.

Gated condition tokens
$Zc = G(Z) = M \odot Z$
Only conditional subspace is activated.

LoRA adaptation
Efficient injection without full fine-tuning.

Attention modulation
Q'c, K'c, V'c = pretrained projections + λF
Text/noise projections remain unchanged.

Design principle

- Visual region controls appearance and structure.
- Gating prevents unintended drift in text/noise tokens.
- LoRA keeps training lightweight while enabling object-level reconstruction.

What it solves

- Fills hidden object regions coherently.
- Maintains texture/style consistency with visible parts.
- Avoids over-reliance on ambiguous text prompts.

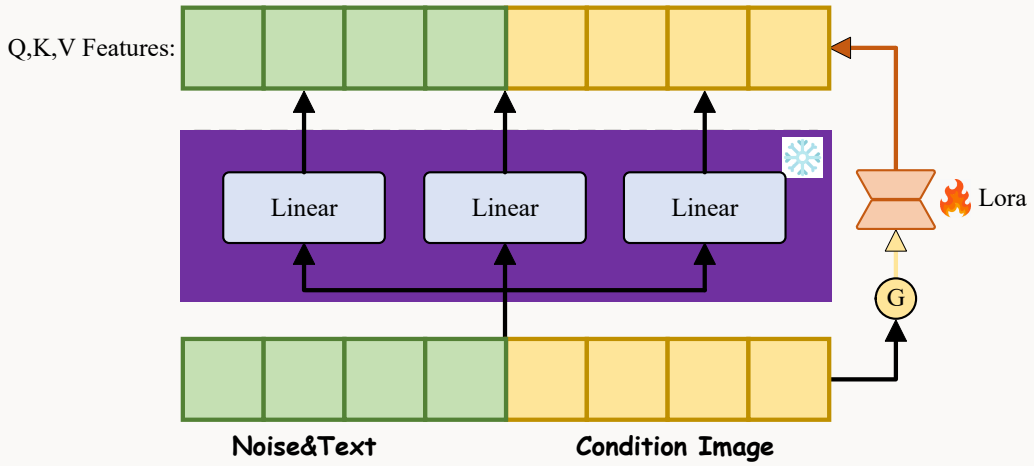


Fig. 3: Condition-Gated LoRA

Stage 3: Color-Invariant Transparent Decoder

METHOD 3/3

PRISM estimates alpha while preserving the generated RGB values.

1

Identify gray background

Mean intensity + RGB variance + saturation constraints

2

Repair mask topology

Morphological closing/opening removes holes and isolated noise

3

Smooth only boundaries

Gaussian filtering is localized to the edge transition region

4

Output RGBA layer

O = [R, G, B, A'] with color integrity preserved

Decoder advantage
Separates transparency refinement from color synthesis: alpha improves without changing object colors.

Experimental Protocol

Evaluation emphasizes visible-region fidelity and generative quality of occluded areas.

Dataset
PrismLayersPro processed for PRISM
Training: 1,985 images → 8,321 element layers
Test: 181 images → 819 element layers

Training
FLUX.1-dev backbone
AdamW, lr 1×10^{-4}
LoRA rank 128; frozen backbone
20 Euler denoising steps

Metrics
PSNR / SSIM: fidelity
LPIPS: perceptual similarity
FID: distributional realism

Evaluation focus

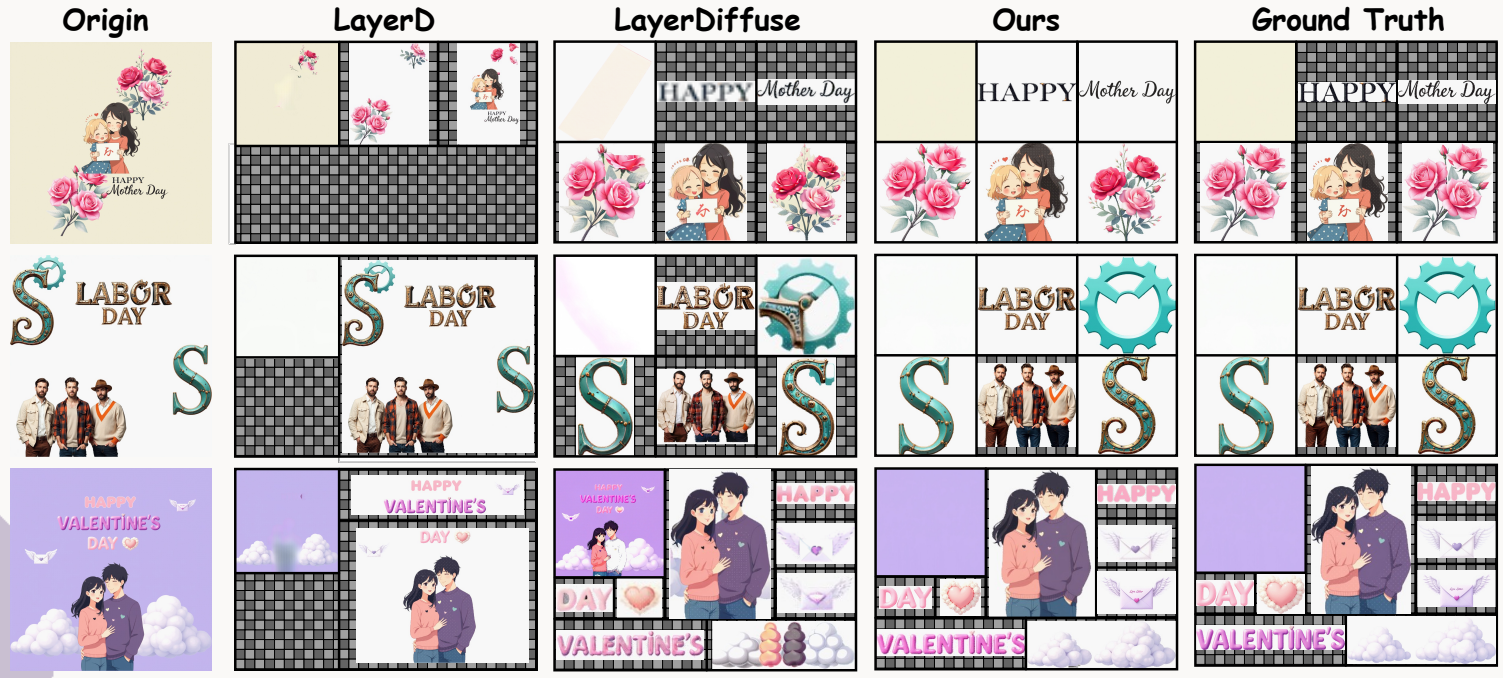
- Complete object layers are required even when original elements are occluded.
- Alpha ground truth for hidden regions is not directly observable; therefore fidelity and realism are both measured.
- Baselines: LayerD and LayerDiffuse under standardized input/output settings.

Results: Accuracy, Realism, and Ablation Evidence

RESULTS

PRISM substantially improves element-layer reconstruction and maintains the best quality-efficiency trade-off.

Method	Element PSNR ↑	Element SSIM ↑	Element LPIPS ↓	Element FID ↓
LayerDiffuse	16.45	0.8098	0.2741	90.74
LayerD	14.22	0.6522	0.6437	142.2
PRISM	**30.35**	**0.9612**	**0.0606**	**24.33**



Ablation finding
Gate variant: SSIM 0.97 / PSNR 31.48 / FID 19.66
512^2 gives best quality-efficiency trade-off.

Against LayerD
+16.13 dB element PSNR
90.6% lower LPIPS

Against LayerDiffuse
+13.90 dB element PSNR
3.7× lower element FID

Visual Results & Take-home Message

Fig. 5: detected element regions become reconstructed RGBA layers with synthesized occlusions and refined alpha

Complete layers
Occluded structures are synthesized to maintain object integrity.

Coherent appearance
Visible texture and style guide reconstruction through Condition-Gated LoRA.

Editing-ready output
RGBA layers have smooth alpha and preserve RGB color values.

PRISM: detect each element, reconstruct what is missing, and decode it into a usable transparent layer.