# Gen4Gen: Generative Data Pipeline for Generative Multi-Concept Composition

BMVC 2025

Our Project Page ↓

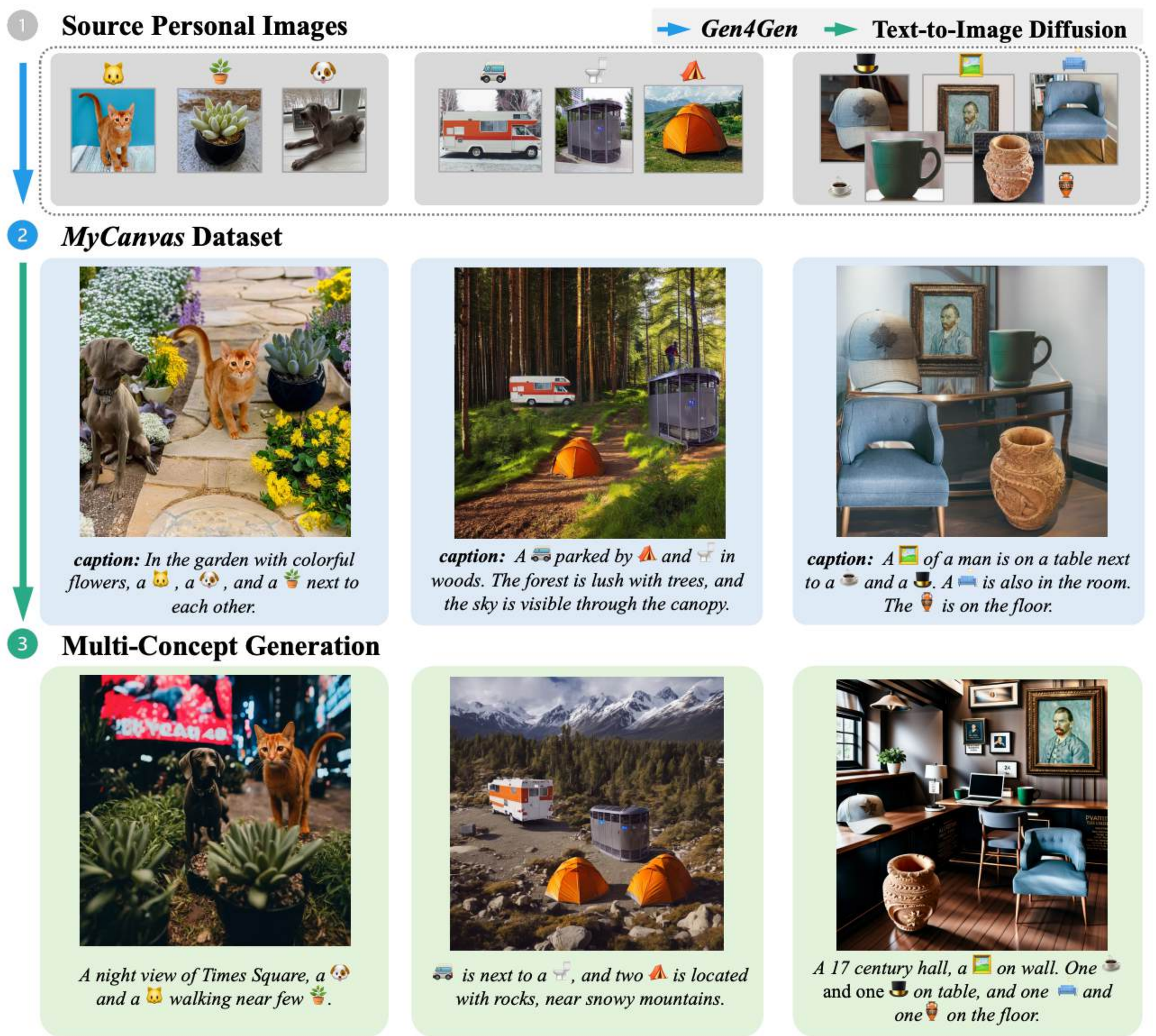

Chun-Hsiao Yeh[1*], Ta-Ying Cheng[2*], He-Yen Hsieh[3*], David Chuan-En Lin[4], Yi Ma[1,5], Andrew Markham[2], Niki Trigoni[2], H.T Kung[3], and Yubei Chen[6]

[1]UC Berkeley [2]University of Oxford [3]Harvard University [4]CMU [5]HKU [6]UC Davis


## Introduction & Motivation

### Failures of Multi-Concept Personalization


### Current Challenges & Our Contributions




**Challenges:**

→ Stable Diffusion Models are largely learned from LAION dataset

→ LAION data often suffer from weak text-image alignment

→ Hard for diffusion models to generate complex and multi-concept images

**Contributions:**

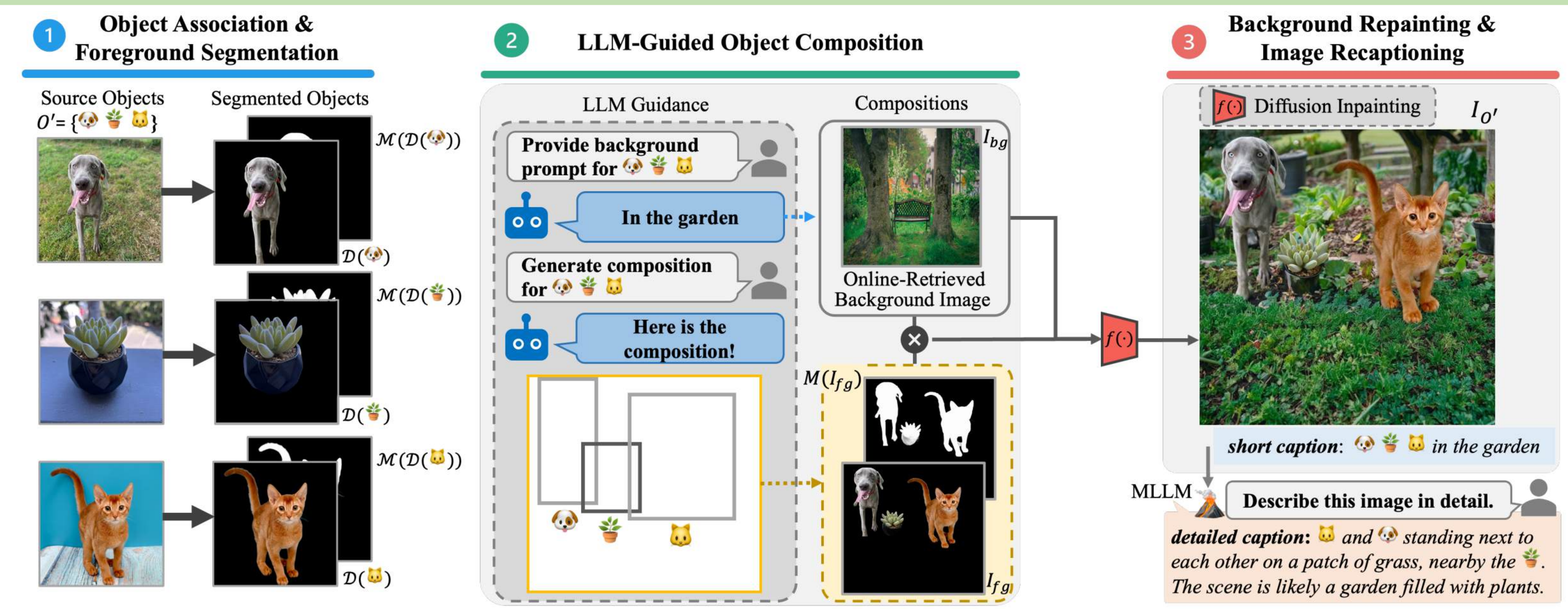
→ from model-centric solutions to a data-driven approach

→ We propose *Gen4Gen*, a generative data pipeline that produces MyCanvas

→ Introduce a high-quality benchmark dataset with well-aligned image-text pairs

## Methodology

### How Do We Create Multi-Concept Composition?




(1) **Foreground Segmentation:** A saliency detector is used segment foreground objects

(2) **LLM-Guided Composition:** A LLM is prompted to generate background and bounding box layouts

(3) **Repainting and Recaptioning:** A diffusion inpainting places the objects into a background image

### My Canvas Dataset Summary


- MyCanvas is a benchmark dataset for multi-concept personalization, created by the Gen4Gen pipeline
- We collected **150 objects** and created **41 compositions**
- Our final benchmark are around high-quality 3000 samples selected from over 10K initial generations

## Quantitative and Qualitative Evaluations

### Comparison with Multi-Concept Baselines


### Quantitative Results and Ablation Studies

**Personalization Alignment (CP-CLIP / TI-CLIP)**

| | <= 3 Objects | | 4 Objects | | 5 Objects | |
|---|---|---|---|---|---|---|
| | CP-CLIP↑ | TI-CLIP | CP-CLIP↑ | TI-CLIP | CP-CLIP↑ | TI-CLIP |
| CD + Source Images | 0.26 | 0.16 | 0.21 | 0.13 | 0.23 | 0.17 |
| CD + MyCanvas | 0.41 | 0.17 | 0.47 | 0.17 | 0.50 | 0.15 |
| Ours + MyCanvas | **0.51** | 0.17 | **0.55** | 0.16 | **0.57** | 0.14 |
| vs. baseline | +0.25 | - | +0.34 | - | +0.34 | - |

**User Preference Study (Score: 1 to 10)**

| | Prompting(Ours) + MyCanvas | | CD + MyCanvas | | CD + Source Images | |
|---|---|---|---|---|---|---|
| | Image Alignment | Text Alignment | Image Alignment | Text Alignment | Image Alignment | Text Alignment |
| <= 3 Concepts | **8.4** (± 1.1) | **8.2** (± 2.6) | **6.8** (± 1.6) | **7.1** (± 2.1) | **3.5** (± 1.8) | **3.9** (± 2.6) |
| 4 Concepts | **8.6** (± 1.4) | **8.9** (± 2.7) | **7.3** (± 1.8) | **7.5** (± 2.7) | **3.7** (± 1.9) | **3.4** (± 2.8) |
| 5 Concepts | **8.8** (± 1.1) | **8.6** (± 2.7) | **6.6** (± 1.6) | **6.4** (± 2.5) | **3.9** (± 1.8) | **3.3** (± 2.7) |

**Quality Evaluation of MyCanvas (Rank: 1 to 5)**

| | Rank: 1 | Rank: 2 | Rank: 3 | Rank: 4 | Rank: 5 | Total Images |
|---|---|---|---|---|---|---|
| <= 3 Concepts | 9 (3.4 %) | 43 (16.3 %) | 72 (27.3 %) | **84 (31.8 %)** | **56 (21.2 %)** | **264** |
| 4 Concepts | 16 (6.0 %) | 53 (19.8 %) | 112 (42.0 %) | **54 (20.2 %)** | **32 (12.0 %)** | **267** |
| 5 Concepts | 19 (7.1 %) | 63 (23.4 %) | 127 (47.2 %) | **42 (15.6 %)** | **18 (6.7 %)** | **269** |