



# SketchDeco:

Training-Free Latent Composition for Precise Sketch Colourisation

Chaitat Utintu Yi-Zhe Song  
SketchX, CVSSP, University of Surrey, UK

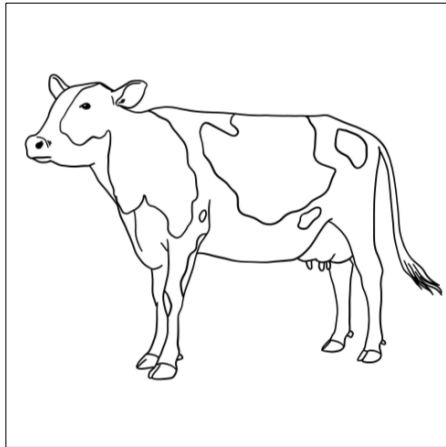


sample results from the proposed method



Pick a region. Pick a colour. Fill it in.

Easy for humans ... surprisingly hard for AI.



"White Cow with large orange patch on the side, turquoise on neck and front legs."

ControlNet



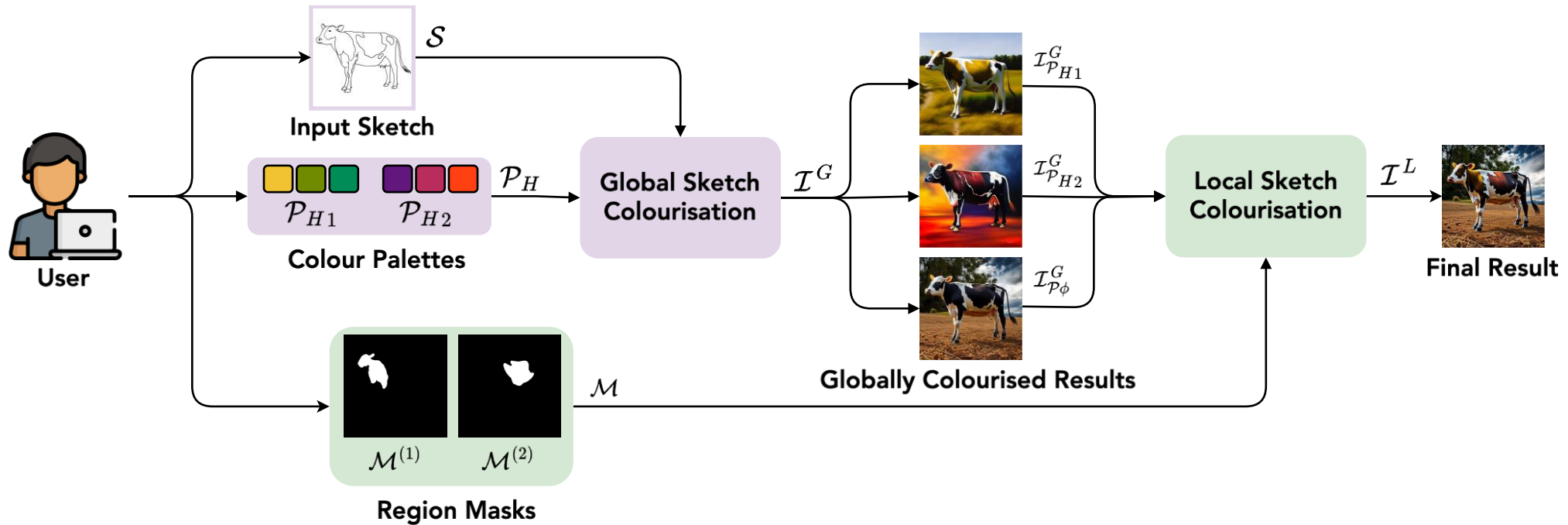
SketchDeco



- ❖ **Task:** colour a sketch; easy for humans, surprisingly hard for AI
- ❖ **Text prompts:** spatially ambiguous, can't guarantee where each colour lands
- ❖ **Trade-off in prior work:** precise control or high quality, never both
- ❖ **Our aim:** precise region and colour control with **no fine-tuning**



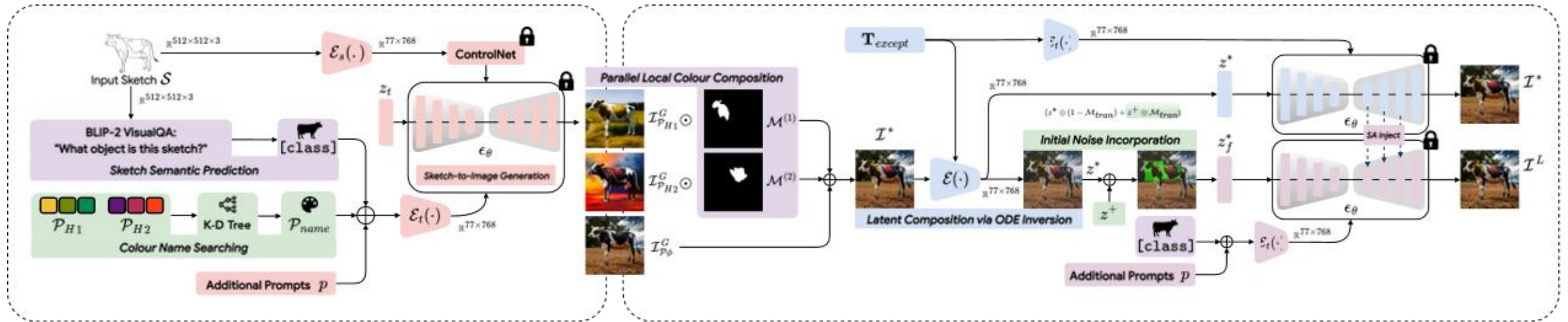
# Our idea: three intuitive inputs from the artist



- ❖ **Sketch:** a black-and-white line drawing
- ❖ **Region masks:** drawn in Photoshop, marking *where* each colour goes
- ❖ **Colour palettes:** hex codes giving *what* colour goes in each region
- ❖ **Output:** a vivid, region-precise image in 15–20 inference steps



# Framework: a divide-and-conquer pipeline



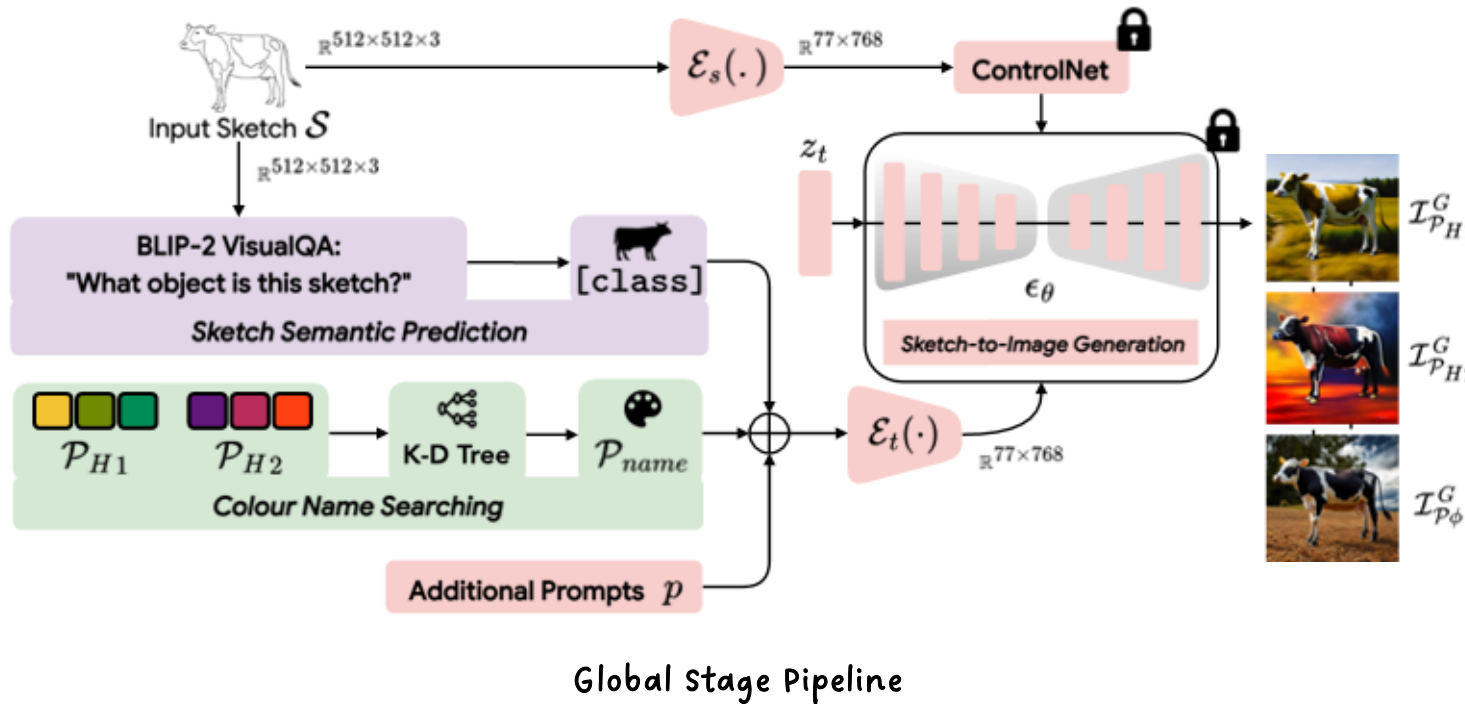
Global Sketch Colourisation Stage

Local Sketch Colourisation Stage

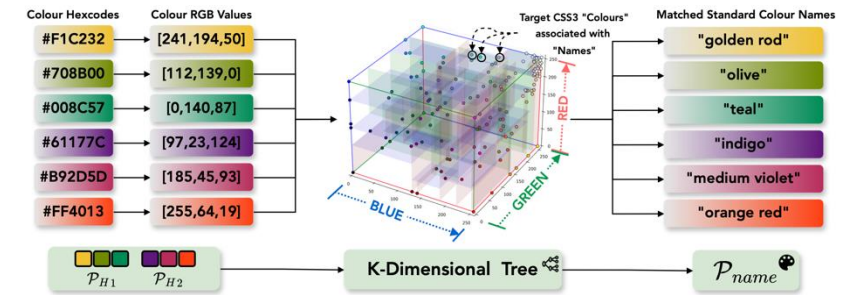
- ❖ **Global stage:** generate plausible, globally-coherent colourisations of the whole sketch
- ❖ **Local stage:** compose the coloured regions onto a background, then re-denoise
- ❖ **Latent composition:** preserves colour fidelity while smoothing region boundaries
- ❖ **Off-the-shelf:** frozen Stable Diffusion and ControlNet, no training



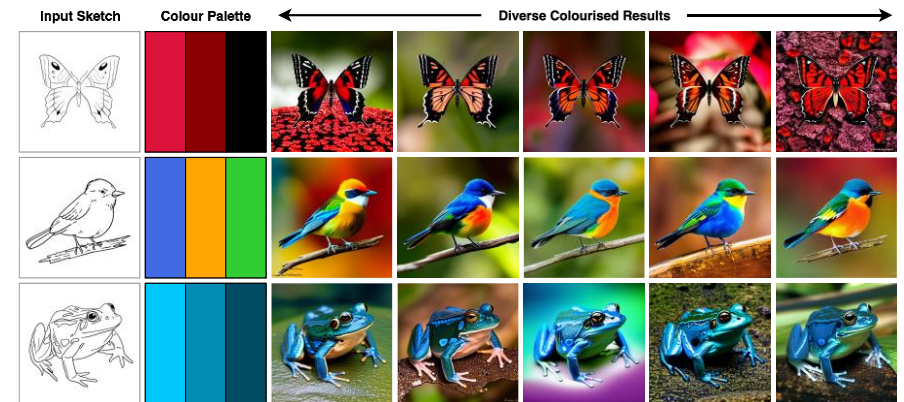
# Stage 1: Global Sketch Colourisation



Global Stage Pipeline



Colour Name Searching with K-D Tree



Diverse Global Colourised Results

- ❖ **BLIP-2 VQA:** asks "what object is this sketch?" to get the class label
- ❖ **K-D Tree:** maps each hex code to its nearest CSS3 named colour (e.g. #F1C232 → "golden rod")
- ❖ **ControlNet + SD:** paints the whole sketch, one image per palette
- ❖ **Background image:** an extra, palette-free render is generated for the local stage

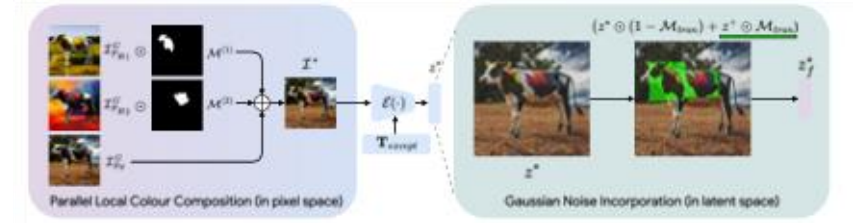
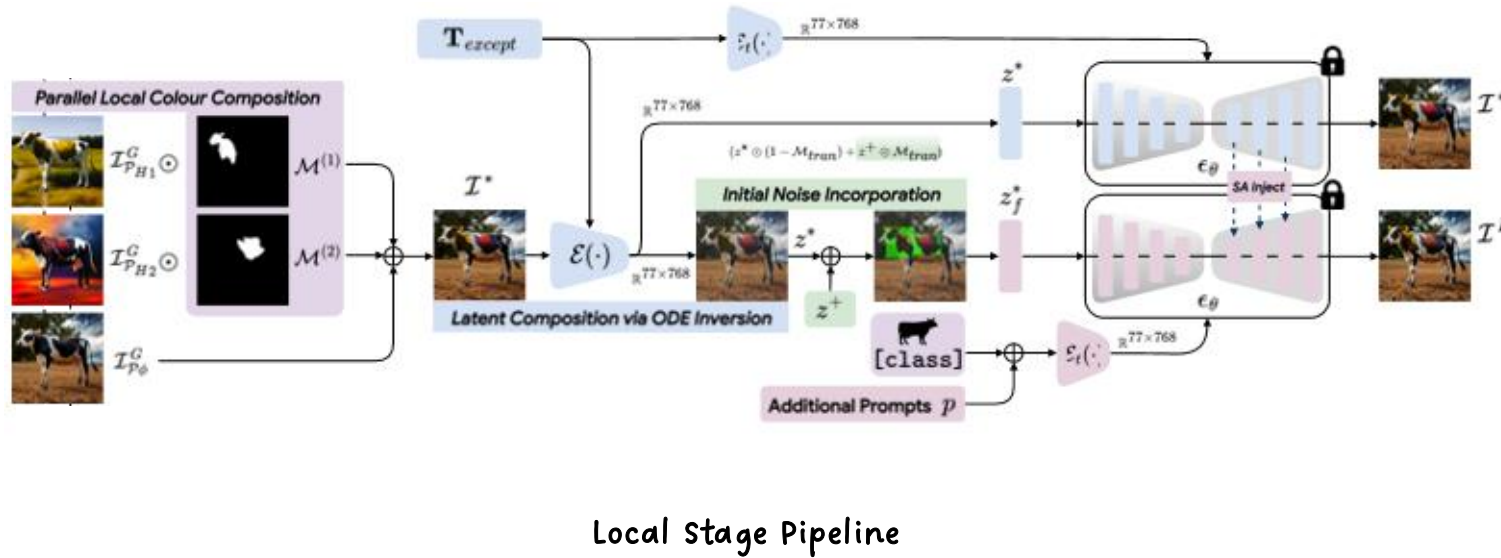




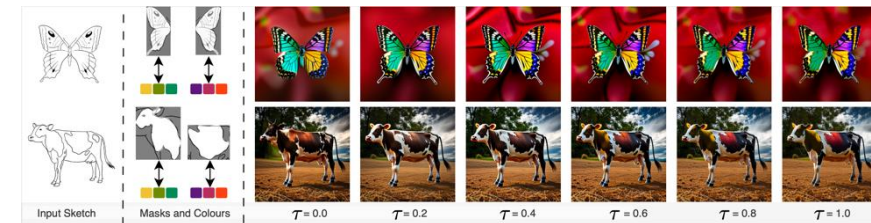
- ❖ **Multiple seeds:** each palette produces several global candidates
- ❖ **Side-by-side preview:** users compare variants in pixel space, not latent space
- ❖ **Pick-and-mix:** choose the best foreground and background combination
- ❖ **Creative control:** ideal for design domains like interior or product mock-ups



# Stage 2: Local Sketch Colourisation



Latent Composition Process



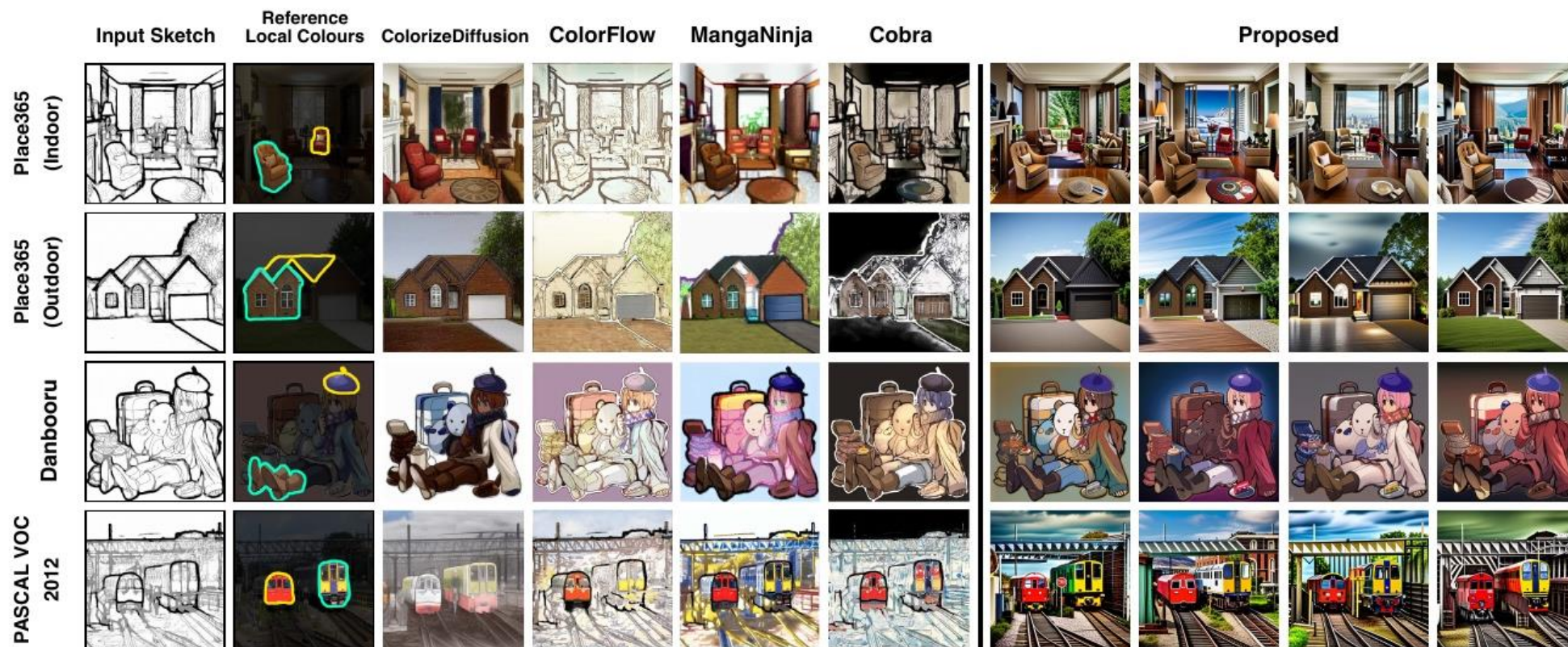
Scaling Factor  $\tau$  Tradeoff

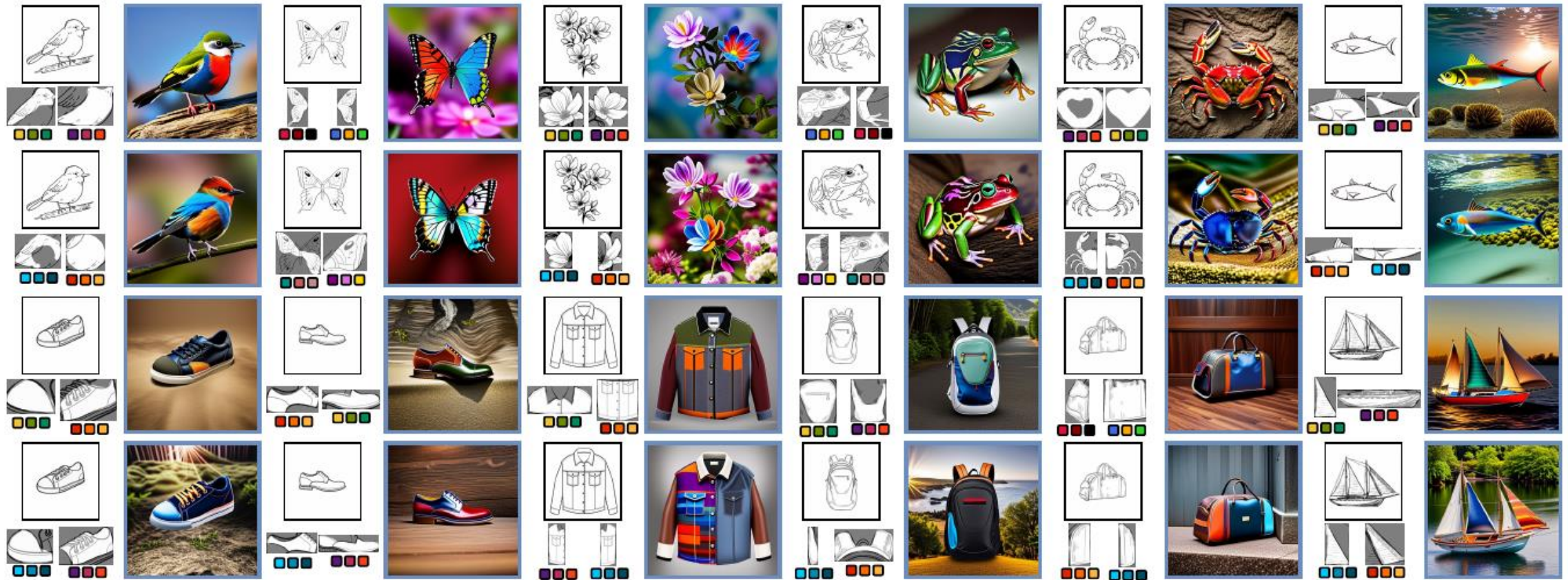
- ❖ **Compose:** paste each masked, pre-coloured region onto the background  $\rightarrow I^*$
- ❖ **Invert:** map  $I^*$  to a latent  $z^*$  using DPM-Solver++ with an exceptional prompt
- ❖ **Noise on seams:** add Gaussian noise only along mask edges, so the model heals the joins
- ❖ **Scaling factor  $\tau$ :** balances colour fidelity vs blending, set to  $\tau = 0.4$



# Evaluation: local sketch colourisation

Methods	Place365 (Indoor)					Place365 (Outdoor)					Danbooru2023					PascalVOC2012				
	FID↓	LPIPS↓	PSNR↑	SSIM↑	DCCW↓	FID↓	LPIPS↓	PSNR↑	SSIM↑	DCCW↓	FID↓	LPIPS↓	PSNR↑	SSIM↑	DCCW↓	FID↓	LPIPS↓	PSNR↑	SSIM↑	DCCW↓
ColorizeDiffusion [76]	151.52	0.645	11.18	0.310	15.30	125.936	0.554	12.04	0.386	9.73	229.44	0.455	9.56	0.485	9.65	110.80	0.517	<b>11.42</b>	0.257	24.37
ColorFlow [86]	354.07	0.643	5.97	0.225	17.05	204.45	0.603	8.13	0.361	16.05	94.87	<b>0.313</b>	<b>10.02</b>	<b>0.541</b>	10.23	367.69	0.491	8.14	0.295	14.98
MangaNinja [42]	134.57	0.548	9.57	0.331	15.19	127.70	0.568	11.24	0.403	9.46	304.17	0.507	9.57	0.485	19.61	289.21	0.492	9.83	0.302	10.61
Cobra [87]	221.38	0.603	8.01	0.186	14.96	193.58	0.732	4.18	0.107	9.05	138.48	0.373	6.27	0.402	13.57	382.70	0.554	5.72	0.164	13.96
<b>Proposed</b>	<b>123.87</b>	<b>0.527</b>	<b>11.65</b>	<b>0.352</b>	<b>11.85</b>	<b>112.48</b>	<b>0.489</b>	<b>12.51</b>	<b>0.408</b>	<b>7.65</b>	<b>134.54</b>	0.450	7.31	0.416	<b>9.23</b>	<b>95.64</b>	<b>0.481</b>	8.97	<b>0.298</b>	<b>8.89</b>





- ❖ **In-the-wild sketches:** sourced from freepik.com, no per-category tuning
- ❖ **Wide coverage:** animals, plants, fashion items, vehicles, scenes
- ❖ **Region-faithful:** colours land exactly where the user's masks specify
- ❖ **Palette-faithful:** the model uses the user's colours, not the real-world ones





People-Centred AI  
UNIVERSITY OF SURREY

**CVSSP**  
Centre for Vision,  
Speech and Signal  
Processing



UNIVERSITY OF  
**SURREY**

**CVPR**  
JUNE 3-7, 2026



**DENVER**  
**COLORADO**

# Thank You for Watching!

## SketchDeco:

### Training-Free Latent Composition for Precise Sketch Colourisation

Chaitat Utintu    Yi-Zhe Song

SketchX, CVSSP, University of Surrey, UK



Project Page

