Exploring Image Representation with Decoupled Classical Visual Descriptors

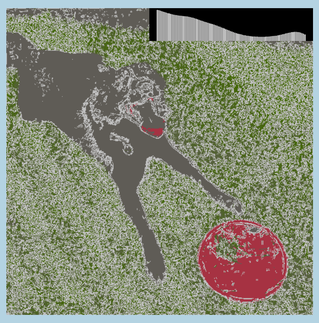
Chenyuan Qu[1], Hao Chen[1,2], Jianbo Jiao[1]

[1] The MIx group, University of Birmingham; [2] University of Cambridge

1. Motivation

Classical visual descriptors are **well-established, deterministic** and **human-interpretable**. Humans can understand the image with incomplete visual concepts. We are interested: “Can deep learning model such an ability?”

Original



Combined Descriptors

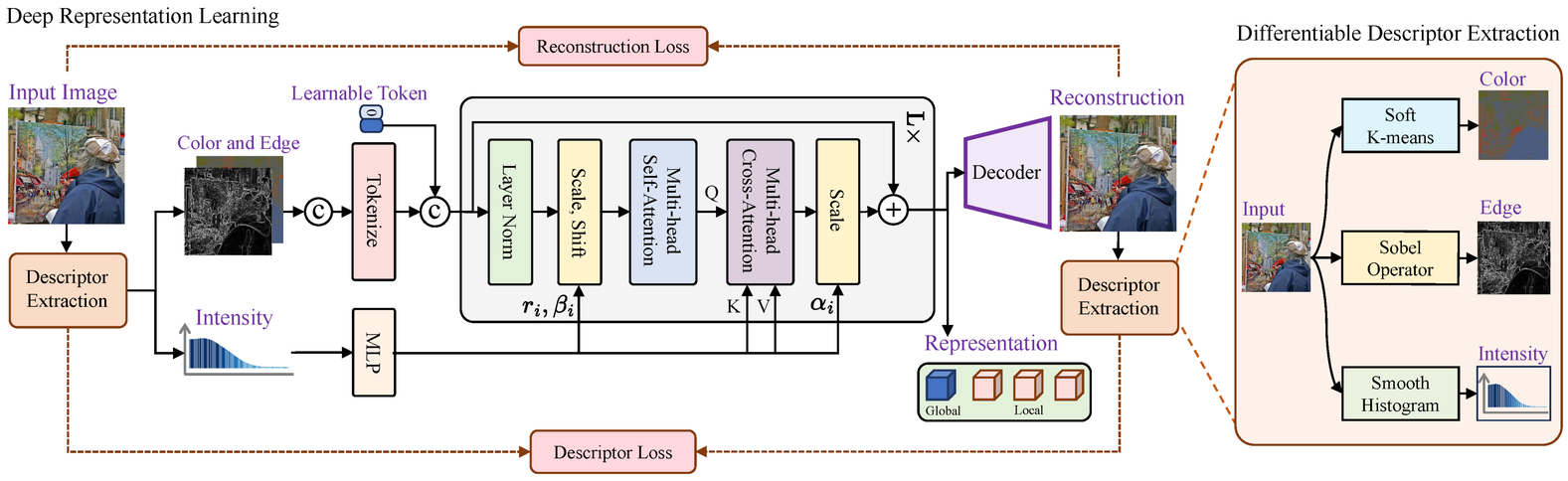
Human Depicted

VisualSplit (Ours)

Answer: *Yes.* We present **VisualSplit**, which reconstructs images from **decoupled descriptors** to learn controllable, interpretable representations

2. Methodology

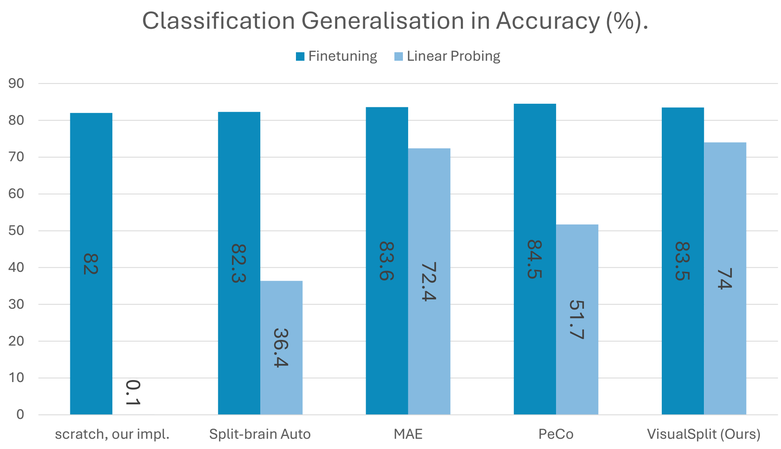
- **Edge map** (Sobel) → local geometry
- **Colour segmentation map** (soft k-means on a/b) → region-wise chroma
- **Grey-level histogram** (100 bins) → global intensity



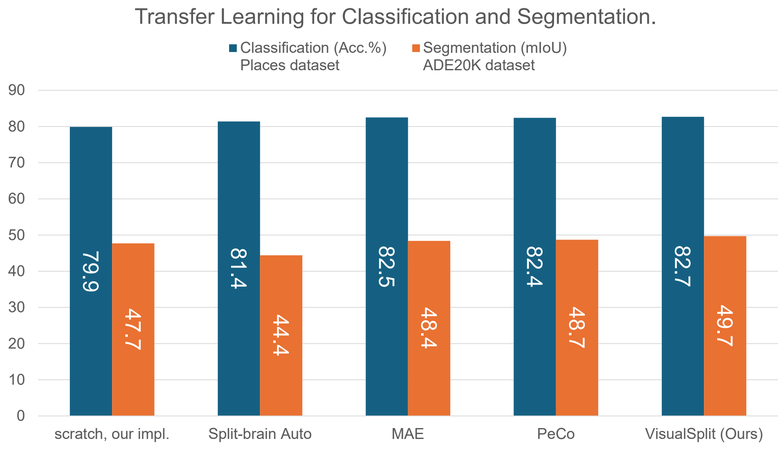
💡 **Takeaway:** Decoupled, sparse inputs act as built-in “missing information,” encouraging robust, structured representations without patch masking.

3. Representation Analysis

Classification. Pre-training on decoupled classical descriptors (edges, colour segments, intensity histogram) yields explicit, transferable features



Segmentation. Transferring to ADE20K with a lightweight SegFormer head.



💡 **Takeaway:** Descriptor-driven pretraining learns structure- and colour-aware features that transfer well.

4. Visual Restoration

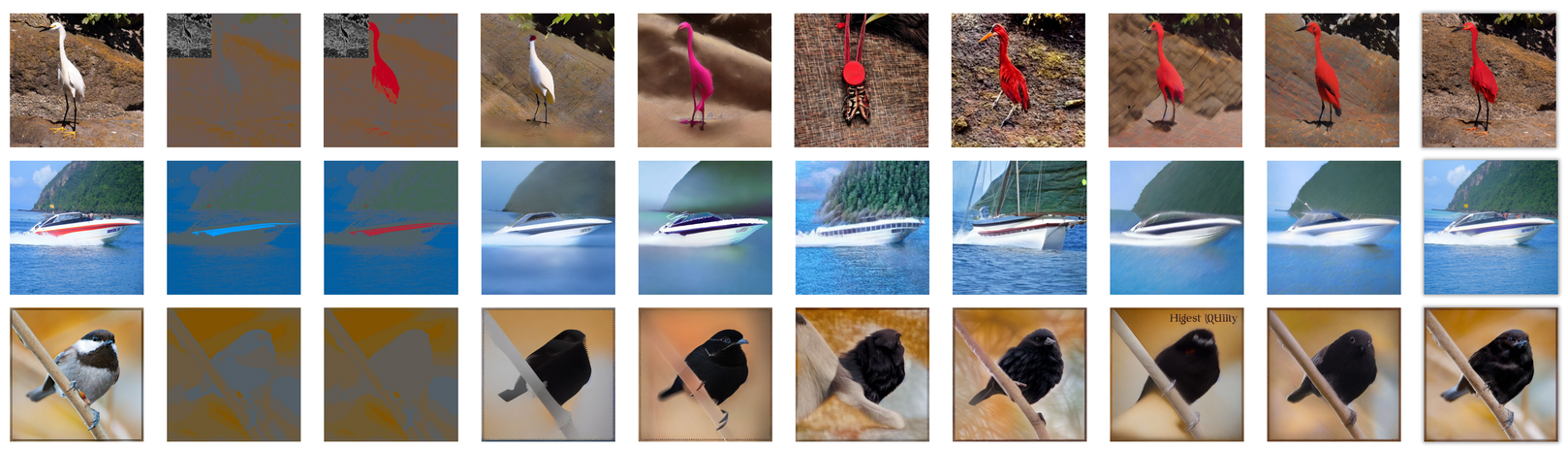
We plug VisualSplit features into Stable Diffusion 1.5 (global guidance via an IP-Adapter-style pathway; local guidance via a ControlNet-style pathway). Using only descriptors as conditions

Original | Descriptors (Input) | w/o prompt / w/ prompt ControlNet | w/o prompt / w/ prompt T2I-Adapter | w/o prompt / w/ prompt ControlNet++ | ***Ours***

💡 **Takeaway:** With descriptors as conditions, our guidance yields sharper, more faithful restorations with fewer artefacts—no prompt required.

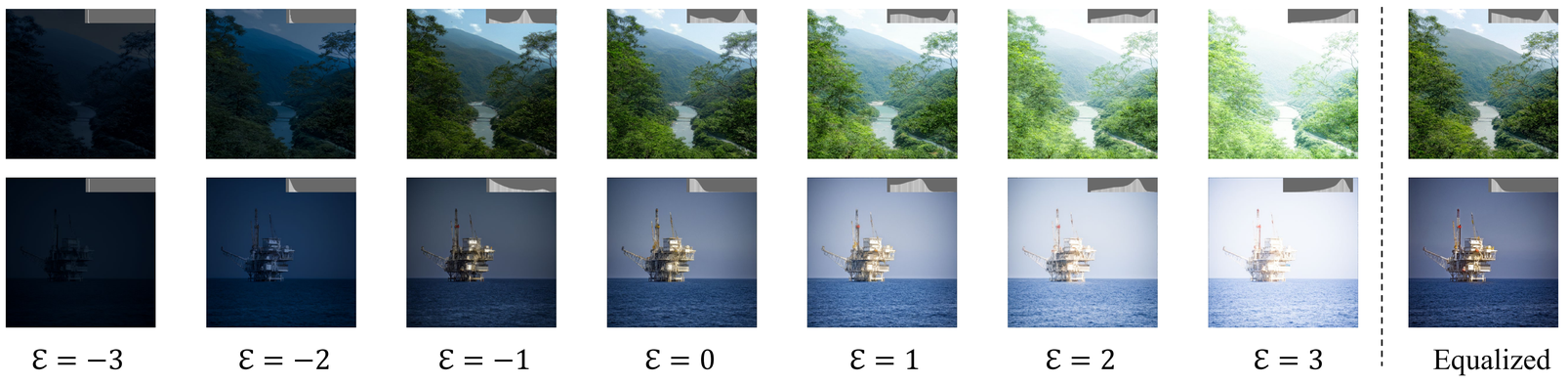
5. Visual Editing

Colour editing. Simply edit the segmented colour map.



Original | Descriptors (Original) | Descriptors (Edited) | w/o prompt / w/ prompt ControlNet | w/o prompt / w/ prompt T2I-Adapter | w/o prompt / w/ prompt ControlNet++ | ***Ours***

Brightness editing. Adjust the grey-level histogram to control exposure/brightness independently of geometry and colour.



$\mathcal{E}=-3$ | $\mathcal{E}=-2$ | $\mathcal{E}=-1$ | $\mathcal{E}=0$ | $\mathcal{E}=1$ | $\mathcal{E}=2$ | $\mathcal{E}=3$ | Equalized

💡 **Takeaway:** Disentangled controls (colour map + grey-level histogram) enable precise, localised, edge-preserving edits that are predictable and repeatable—without re-training or prompts.

Project Page: https://chenyuanqu.com/VisualSplit/

Code: https://github.com/HenryQUQ/VisualSplit