CLICKDIFFUSION

Harnessing LLMs for Interactive Precise Image Editing

Alec Helbling, Seongmin Lee and Polo Chau Georgia Institute of Technology

CLICK DIFFUSION leverages LLMs to combine direct manipulation and natural language instructions, enabling precise image editing.

Direct Manipulation

+ "natural language"

Our approach is:

Precise

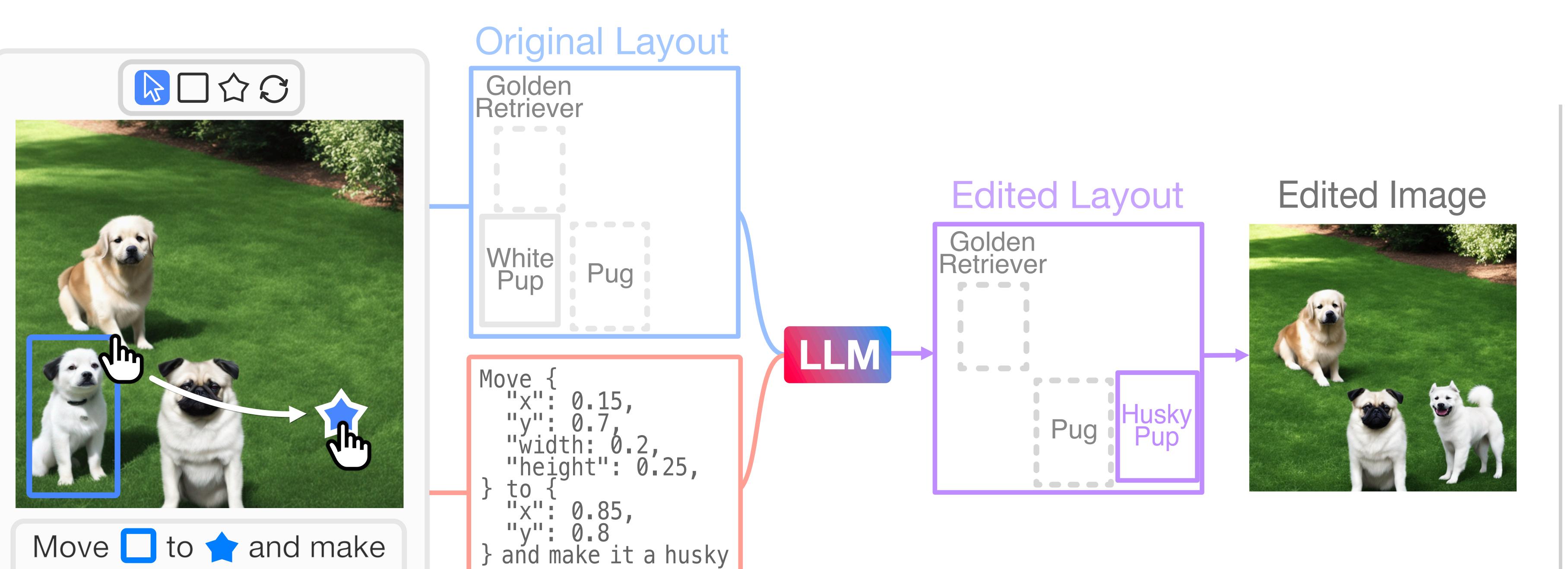




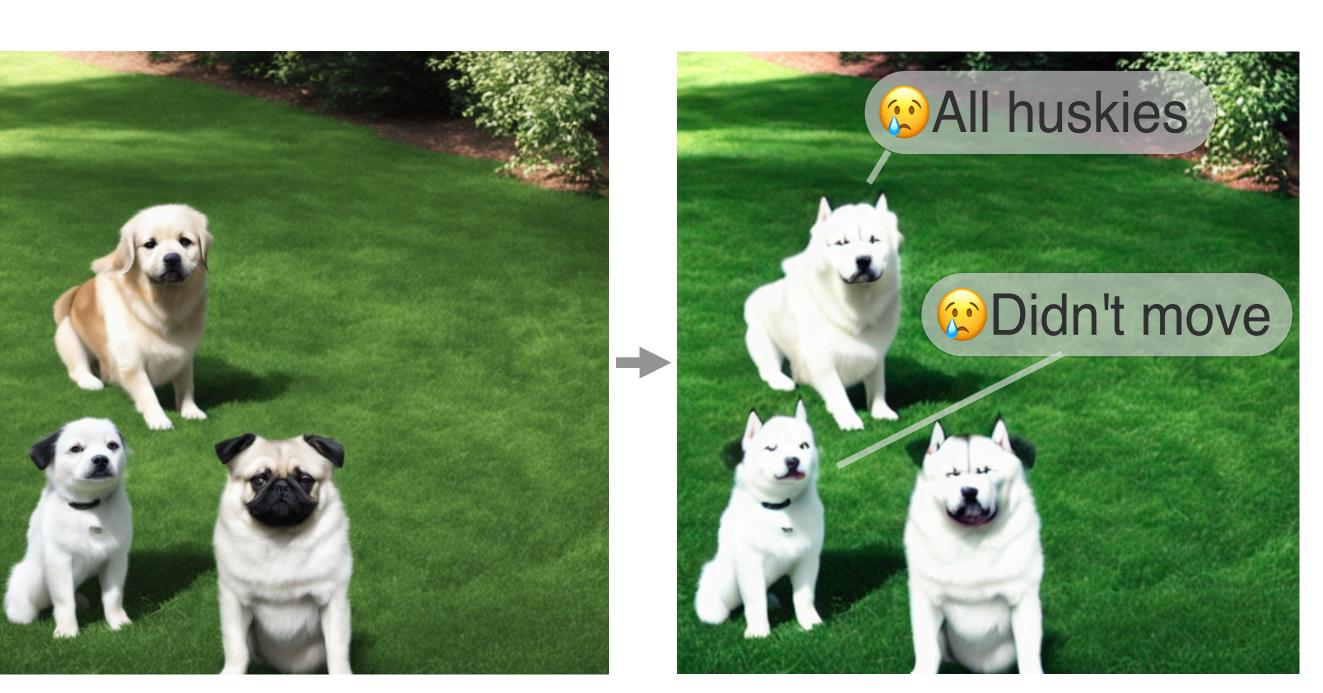
Check out the paper!



Overview of the ClickDiffusion Framework



Comparison to Text Only Editing



Move the dog on the left to the right of the dog in the middle and make it a white husky

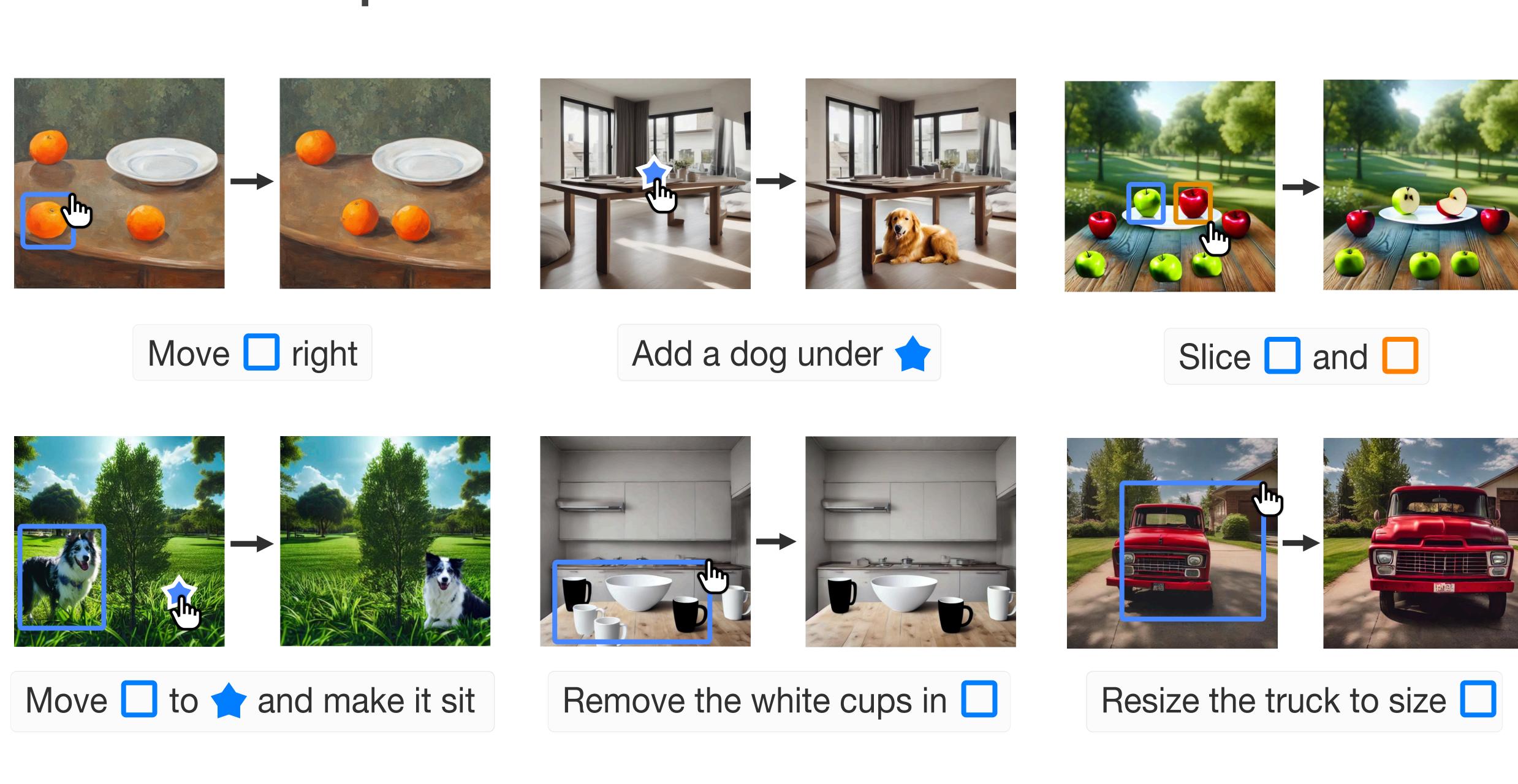
Verbose prompt

Various Edits Enabled by ClickDiffusion

Text Instruction

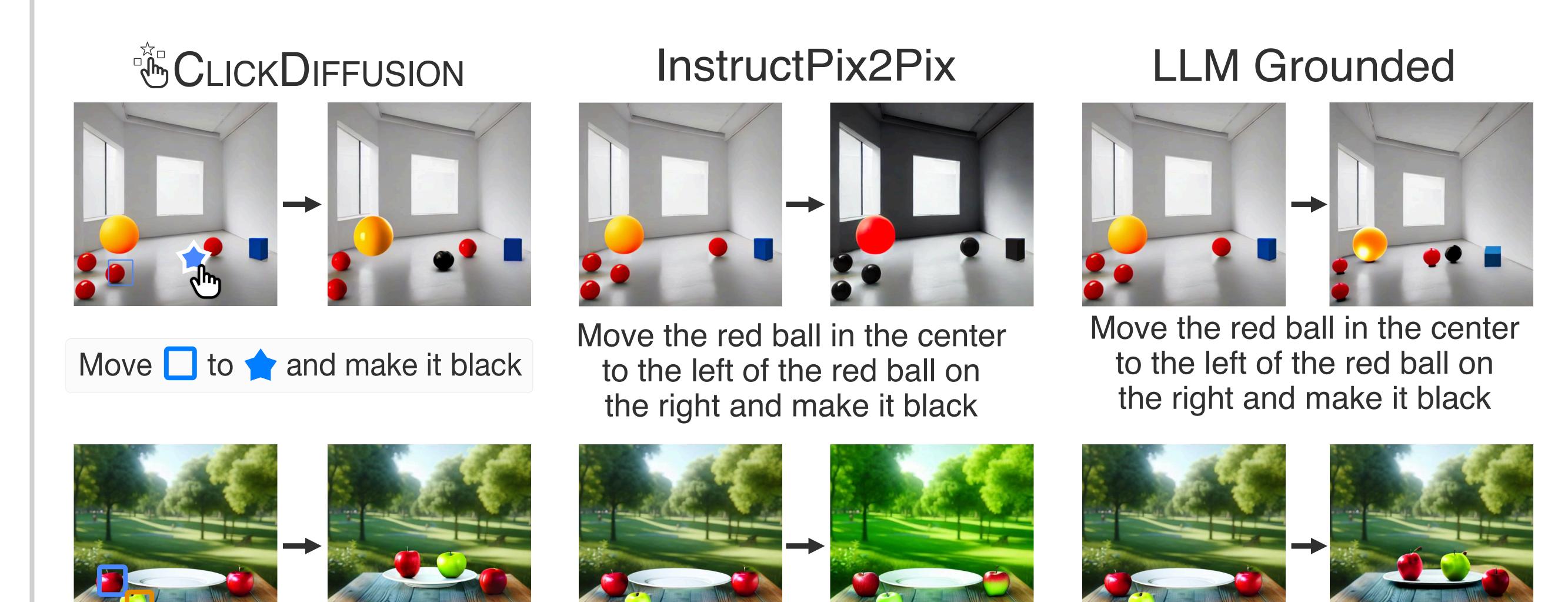
it a husky.

- Our approach enables simple manipulations like moving objects, adding them, and removing them.
- Our approach also excels at specifying manipulations that require a combination of text and spatial information.



Comparison to Other Image Editing Methods

- Our method allows movement and localization of edits unlike many existing methods (e.g. InstructPix2Pix).
- CLICK DIFFUSION requires less complex instructions than text only approaches.



Move the top left red apple and

top green apple onto the plate

Move and to the plate

Move the top left red apple and top green apple onto the plate