## ClickDiffusion

## Harnessing LLMs for Interactive Precise Image Editing

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

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

| Direct |
| :---: |
| Manipulation |${ }^{\text {ing }}+$| ＂natural |
| :---: |
| language＂ |

Our approach is：
$\bigcirc$ Precise $\bigcirc$ Simple $@$ Natural

Check out the paper！


囚ロ凸0


Move $\square$ to and make it a husky．

Overview of the ClickDiffusion Framework

## Original Layout



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
－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．


Move $\square$ right


Move $\square$ to and make it sit


Add a dog under



Slice $\square$ and $\square$


Comparison to Other Image Editing Methods
－Our method allows movement and localization of edits unlike many existing methods（e．g．InstructPix2Pix）．
－ClickDiffusion requires less complex instructions than text only approaches．


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

