Edit me: A Corpus and a Framework for Understanding Natural Language Image Editing

Ramesh Manuvinakurike*[2] , Jacqueline Brixey*[2], Trung Bui[1] , Walter Chang[1] , Doo Soon Kim[1] , Ron Artstein[2] , Kallirroi Georgila[2]

Image Edit Requests

Examples from Reddit forum:

There is a spot on my wedding dress. Can someone please remove it. Please!

Can you please fix the glare on my dog's eyes? I lost him today and he means the world to me.

Can you please remove the people in the background? This is the only surviving photo of my mom and I would like to preserve it.

Goal: photo editing software that interprets and executes requests

Corpus

334 images from Visual Genome corpus (Krishna et al., 2015)

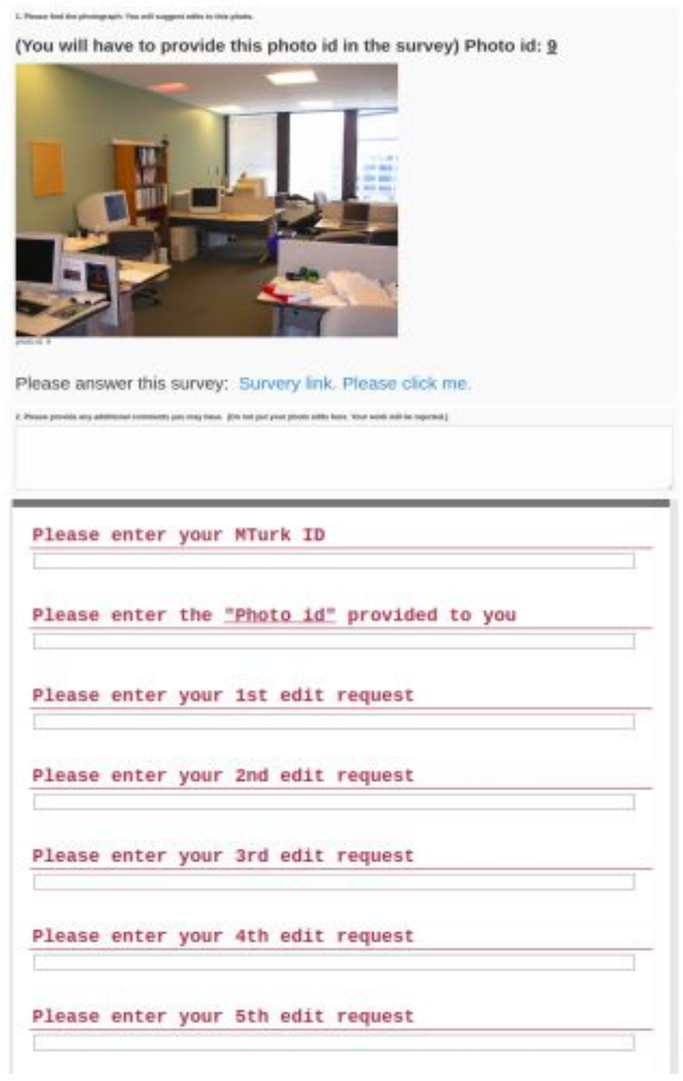
animals, city scenes, food, nature/landscapes, indoor scenes, people, sports, and vehicles

5 Amazon Mechanical Turk users per image

≥5 image edit requests per user

Total: 9101 image edit requests

44727 word tokens
4628 unique word types



Language of Image Edits

Different vocabulary for same intent

Crop the window
Cut out the window
Trim to remove the window

Ambiguous requests

Focus on the cat zoom in or crop?

Wide variety of structure

Also comments.

Different language for experts and novices

I would like to see more character and color to the cobblestone sidewalk. It is lovely.

Adjust the brightness on the white tool to avoid making it look plain white.

Annotation Framework

Action Type	Example
Adjust (44.89%)	Increase saturation a bit on the elephants.
Delete (13.70%)	Remove the jacket hanging from the girl's side.
Crop (6.89%)	Crop the photo to eliminate the space to the left and right of the elephants.
Add (6.85%)	Insert a ball hitting the tennis racket.
Replace (2.47%)	Please change the pamphlet she is holding into a dictionary.
Apply (1.44%)	Add a Gaussian blur to the background.
Zoom (0.87%)	Zoom in on the man.
Rotate (0.71%)	The photo looks tilted. Rotate it clockwise so the lines are straight.
Transform (0.62%)	Flip the photo horizontally.
Move (0.60%)	Move the white framed picture to the blue wall.
Clone (0.33%)	Use a cloning tool to blend grass to cover any patches of dirt on the ground.
Select (0.19%)	Select the white dog.
Swap (0.14%)	Please perform a face swap using the man in the yellow shirt and the man in the blue/black polo.
Undo (0.02%)	If possible uncrop photo to allow more space to frame, rather than cut off the bike.
Merge (0.02%)	Blend the grey smudges so they are the same color as the rest of the dirt.
Redo (0.01%)	Redo all white traffic lines in street.
Other (0.01%)	Resize photo to show large elephant and trainer.

Attribute	Properties of the image to adjust, such as saturation
Object	Item to be inserted or deleted.
Region	Location within the image where an action is being applied.
Modifier /Value	Degree or direction of the change
Intention	end goal for the change

Conclusion & Future Work

- Contribution of a novel dataset for natural language image editing.
- Annotation framework defined with reasonable efficiency (Please refer the paper for analysis and numbers).
- Build models for language understanding and extend framework for dialogue interactions.

Adobe

USC Institute for Creative Technologies