Moving Objects in a Scene

Objective

Given an image containing multiple objects the task is to :

1. Segment an object, given via a text prompt
2. Move the object in the scene based on the given horizontal and vertical displacements

Approach

Task 1

The object segmentation task is carried out in a two stage process.
First we use CLIPSeg [1], which takes as input an **image** and a **text prompt** and returns a **confidence map** indicating the presence of the the object named in the text prompt in each pixel via a **Pixel Score**. We normalize the scores to bring it in the range of [0,1] and select the co-ordinates of **one pixel** with a score above a threshold **(0.99 in our case)**
In second stage, we pass the image and the co-ordinates of selected pixel to SAM[2]. SAM returns multiple segmentation maps in which objects on which the input co-ordinates lie our segmented. **We choose the segmentation map with largest area as our final output**. The rational for this is that as we select a pixel, with high probability to be on the object of interest and pass it along with the image on to SAM, it returns multiple masks all of which try to segment out parts of image with distinct boundaries and contain the input pixel in it. By choosing the mask with maximum area we ensure we choosing the complete object and not just parts of it, as certain parts of object which have distinct boundaries are also returned as masks.This is also relevant for the given use case as in these applications we intend to segment and thus displace complete object in a scene and not just parts of the object. We also store the segmentation mask which will be useful in Task 2

Task2

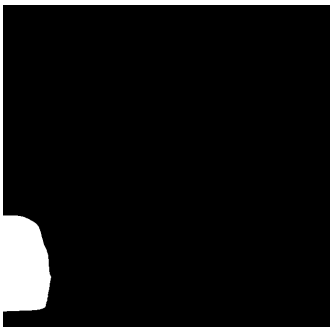
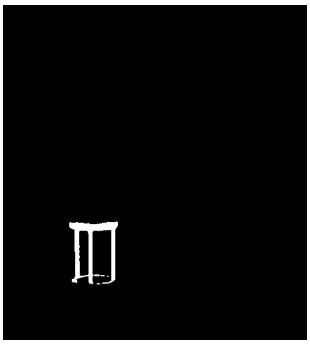
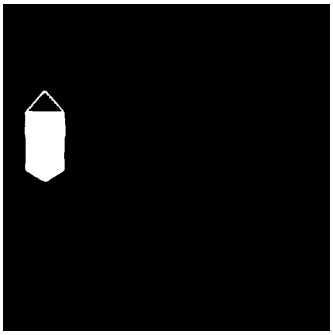
For second task we use DragonDiffusion[3], which takes as input the image to be edited, name of the object to be moved, displacement values x and y axis and also the segmentation mask from Task 1 and moves the object in the scene.

Visual Results

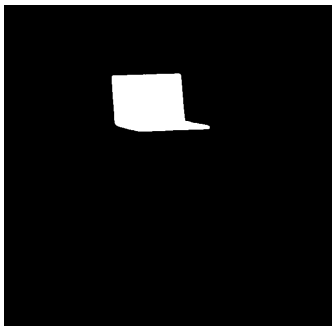
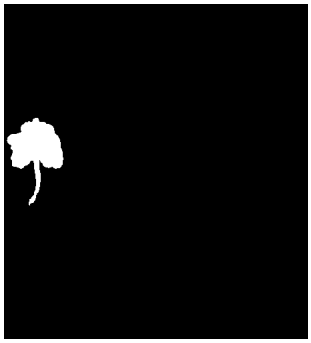
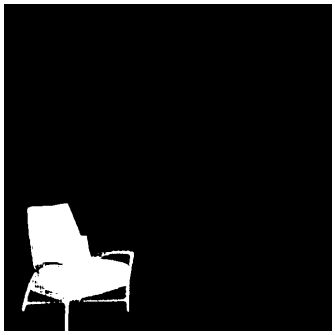
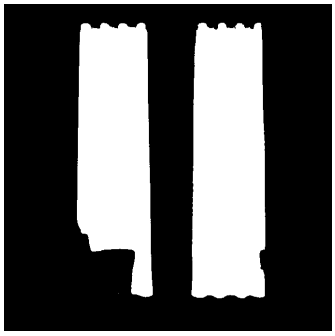
Task 1

Following are the results for Task 1 i.e object segmentation for the 3 given images :

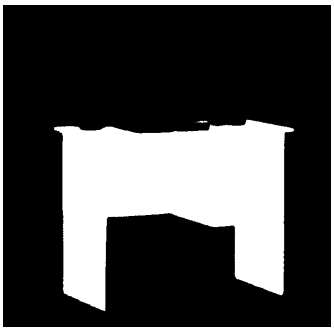
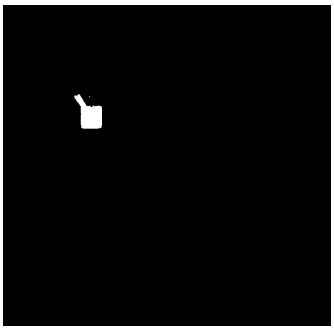
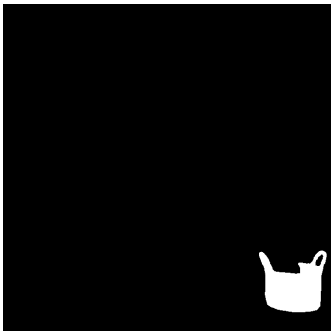
- Table 1 has the required set of segmentation based on the problem statement
- Table 2 and Table 3 has some more segmentation results

Text Prompt	Image	Segmentation Mask	Segmented Image
Bagpack			
Stool			
Wall Hanging			

Tabel 1: Result on Given Images for Task 1

Text Prompt	Image	Segmentation Mask	Segmented Image
Laptop			
Flowers			
Chair			
Curtain			

Tabel 2: Some Extra Results for Task 1

Text Prompt	Image	Segmentation Mask	Segmented Image
Table			
Pencil Stand			
Basket			

Tabel 3: Extra results for Task 1

Task 2

Following are the results on Task 2 in Table 4

Text Prompt	Image	Displacement	Displaced Object
Laptop		x = -30 y = 60	
Chair		x = -50 y = 0	
Bagpack		x = 0 y = 60	
Pencil Stand		x = -50 y = 50	
Wall Hangning		x = 0 y = 80	

Tabel 4: Some Results for Task 2

Strengths

1. With respect to Task 1, i.e object segmentation, the method proved to be robust on the given set of images when tried on various objects in each image
2. In the case of Task 2, when objects are moved against a plain background i.e., the displacement parameters are such that the object will not superimpose any other object in foreground, the method performs the edits with minimal changes in displaced object
3. Usage of SAM in the pipeline helps is accurate segmentation of images

Weakness

1. In Task 2, if multiple objects are very close by in the given image like flowers in a vase or some things kept on a table, we are able to segment them individually, but while displacing such objects either nearby objects gets distorted or the object to be moved changes its appearance
2. There also a challenge of performing edits involving superimposing objects. Such edits usually lead distortion in the object appearance.
3. Also in some case edits with very large displacements, i.e. large values for x and y tend to distort the object.

Failure Cases

Few failure cases have been show in Table 5

Scope for Improvement

1. For Task 2, in the case where we want to move an object in front of another object i.e. moving in unclear background, using a 3D model can be beneficial. In a 2D model like above leads to distortion in both the object in front and at the back.
2. This edit can be done in 3D space and the scene can be rendered back to get the 2D edited image from the desired view.
3. For dealing with large displacements we can follow a iterative process where at each step we do only limited displacements. So first we segment the object,generate the mask and displace it a little bit using Dragon Diffusion (some quantized units in one direction) to get a new image and then again segment the new image and so on in a cyclic manner.

Text Prompt	Image	Displacement	Displaced Object
Stool		x = -50 y = 0	
Bag		x = 0 y = 40	
Vase		x = 40 y = 0	
Bag		x = 40 y = 0	
Flowers		x = 50 y = 50	

Tabel 5: Some Failure Cases for Task 2