# Seamless Object Transfer Between Images

Batuhan Solmaz

January 10, 2025

**Abstract**

This study introduces a method for seamlessly transferring objects between images using simple user prompts. The approach combines the Segment Anything Model (SAM) for extracting objects with Deep Image Blending to ensure they fit naturally into new scenes. SAM uses advanced computer vision techniques to create accurate object masks, while Deep Image Blending adjusts the object's appearance to match the target image by optimizing various aspects like gradients, content, style, and color distribution. The experiments show that this method works well in different scenarios, producing visually realistic and contextually appropriate results. This approach has the potential to improve image editing workflows in areas such as media production, augmented reality, visual storytelling, and beyond.

## 1 Introduction

Image editing is a key part of modern digital media, with applications in areas like graphic design, augmented reality, and personalized content creation. One of the biggest challenges in this field is transferring objects between images while maintaining a realistic look and feel. Traditional methods often struggle with accurately segmenting objects or blending them into new scenes, which can lead to unnatural and visually unappealing results.

This study tackles these challenges by combining the Segment Anything Model (SAM) and Deep Image Blending. SAM is a powerful tool for creating precise object masks based on user prompts, while Deep Image Blending ensures that objects fit naturally into their new environments by optimizing several factors like color, texture, and smooth transitions. Together, these methods produce high-quality results that can be applied to creative design, content creation, and beyond.

The structure of this paper is as follows: Section 2 covers the background and related works, Section 3 defines the problem and system description, Section 4 explains the solution technique, Section 5 presents experiments and results, and Section 6 concludes with a discussion and future directions.

## 2 Background and Related Works

Transferring objects between images is a common task in photo editing, but it is usually done manually using tools like Photoshop. This can be time-consuming and requires a lot of skill to make the final result look natural.

Recent advancements in computer vision have made it easier to automate parts of this process. The Segment Anything Model (SAM) is a tool that helps users extract objects from images by generating precise masks. It works based on prompts like text descriptions or selected points in the image [1].

Another method, called Deep Image Blending, focuses on blending objects into new images. It uses advanced techniques to ensure that the object matches the style, lighting, and texture of the target image. By carefully optimizing different aspects like smoothness and color, it can produce realistic results [2].

Although there aren't any known tools that fully automate transferring objects between images, this project combines SAM for object extraction and Deep Image Blending for integration. This approach allows for seamless object transfer with minimal user effort.

## 3 Problem Definition / System Description

Moving objects from one image to another in a realistic way is a challenging task. The object must be cut out cleanly and blended into the new image so it looks like it belongs there. The main goals of this project are:

- Extract the desired object from a source image accurately.
- Blend the object into the target image so it matches the lighting, colors, and textures.
- Make the entire process as simple and automated as possible.

The project has two main parts:

1. **Segment Anything Model (SAM):** This tool is used to extract the object from the source image. It can generate very accurate masks based on user input, like points or a description of the object.

2. **Deep Image Blending:** This method blends the extracted object into the target image. It works by carefully adjusting the object to match the new image, so the result looks smooth and realistic.

The combination of these two methods provides a solution for transferring objects between images with minimal manual work.

# 4 Solution Technique

This project uses two main techniques to achieve seamless object transfer between images: the Segment Anything Model (SAM) for object segmentation and Deep Image Blending for harmonizing the transferred object with the target image.

## 4.1 Segment Anything Model (SAM)

The Segment Anything Model (SAM) is designed to generate high-quality object masks based on user input.

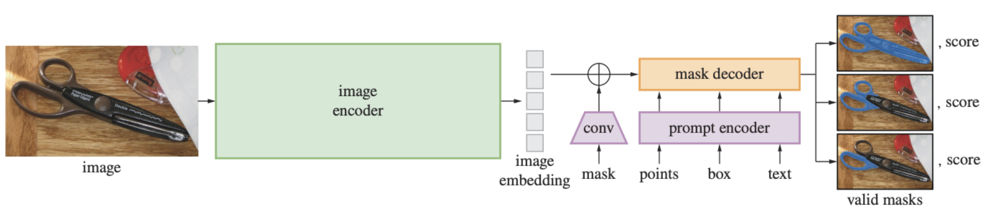


Figure 1: Architecture of the Segment Anything Model (SAM).

SAM has three main components:

- **Image Encoder:** A Vision Transformer (ViT-H) that converts an image into a series of embeddings. These embeddings capture the visual features of the image.
- **Prompt Encoder:** Based on CLIP's encoder, this component processes text descriptions and encodes them into embeddings.
- **Mask Decoder:** This combines the embeddings from the image and prompt encoders to produce accurate segmentation masks. The mask decoder uses a combination of self-attention and cross-attention layers to refine the masks.

SAM is trained using two main loss functions:

- **Focal Loss:**
$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$
This loss helps handle imbalanced datasets by focusing more on hard-to-classify pixels.
- **Dice Loss:**
$$D = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}$$
This loss ensures better overlap between the predicted mask and the ground truth.

## 4.2 Deep Image Blending

Once the object is segmented using SAM, it is blended into the target image using Deep Image Blending. This process ensures that the transferred object matches the style, lighting, and texture of the target image.

The blending process has two stages:

1. **Stage 1: Initial Blending.** The object is placed in the target image, and several loss functions are optimized to smooth the edges and adjust the colors. The blended image is computed as:
$$I_B = I_Z \odot M + I_T \odot (1 - M)$$
Here, $I_Z$ is the segmented object, $I_T$ is the target image, and $M$ is the mask.

2. **Stage 2: Refinement.** The blending is further improved by re-optimizing the image, with a higher focus on style and texture. This step ensures the transferred object looks natural.

Key loss functions used in Deep Image Blending include:

- **Gradient Loss:** Ensures smooth transitions by matching the gradients of the blended image with those of the target and source images.
- **Content Loss:** Preserves the structure of the object by comparing deep features from a pre-trained network (e.g., VGG).
- **Style Loss:** Matches the texture and color style using Gram matrices.
- **Histogram Loss:** Aligns the brightness, contrast, and overall texture between the blended image and the target.
- **Total Variation Loss:** Promotes smoothness in the blended image by penalizing large pixel differences.

The total loss for blending is given as:

$$L_{total} = \lambda_{content} L_{content} + \lambda_{style} L_{style} + \lambda_{hist} L_{hist} + \lambda_{tv} L_{tv}$$

Including these stages ensures the final result is visually coherent and free of artifacts.

# 5 Experiments and Results

In this section, there are six examples to evaluate the performance of the proposed method. Each example consists of a source image, a target image, and the result after transferring the object. The method successfully integrates the objects into new environments.

### 5.1 Example 1: Boat Transfer

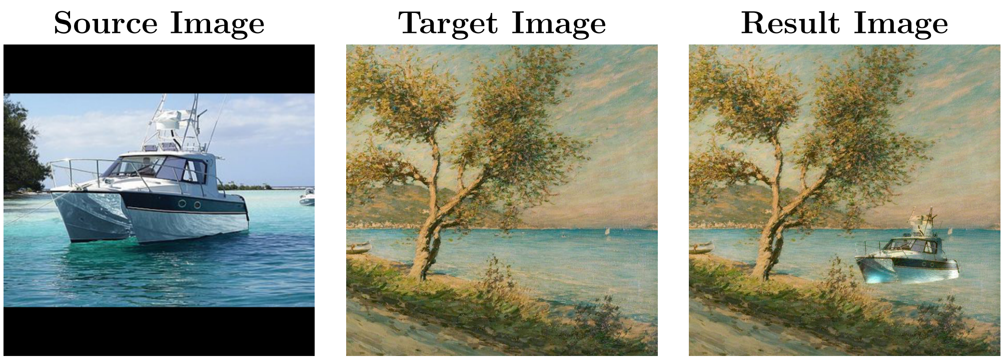


Figure 2: Prompt: "Get boat."

### 5.2 Example 2: Eiffel Tower Transfer

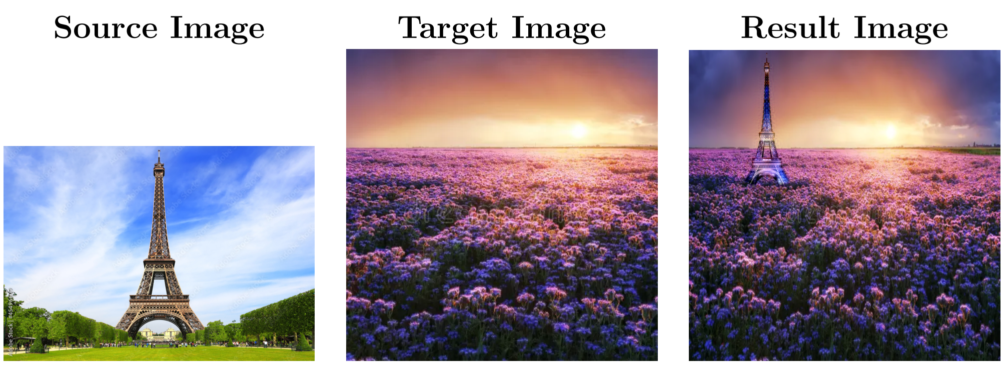


Figure 3: Prompt: "Get Eiffel Tower."

### 5.3 Example 3: People Transfer

**Source Image** **Target Image** **Result Image**

Figure 4: Prompt: "Get all people."

### 5.4 Example 4: Gladiator Transfer

**Source Image** **Target Image** **Result Image**

Figure 5: Prompt: "Get gladiator, sword, and shield."

### 5.5 Example 5: Balloon Transfer

**Source Image** **Target Image** **Result Image**

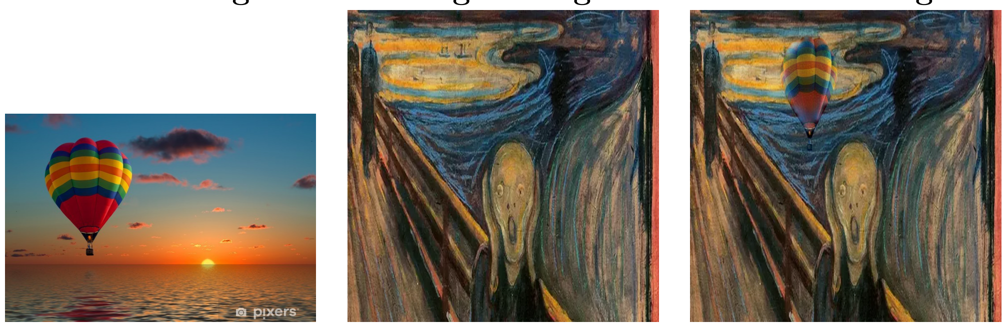


Figure 6: Prompt: "Get balloon."

### 5.6 Example 6: Lion Transfer

**Source Image** **Target Image** **Result Image**


Figure 7: Prompt: "Get the lion on the left."

## 6 Discussion and Limitations

The proposed method demonstrates high-quality object transfer, but it has some limitations:

- **Computational Cost:** The Deep Image Blending process is slow because it involves optimizing multiple loss functions, such as gradient, content, style, and histogram losses, which require significant computation.

- **Edge Artifacts:** In some cases, minor edge artifacts are observed, particularly when source and target images differ greatly in lighting or texture.
- **Prompt Dependency:** The method relies heavily on accurate user prompts for successful object segmentation.

# 7 Conclusion and Future Work

This study combines SAM and Deep Image Blending to achieve seamless object transfer between images. While the results are promising, limitations like slow blending and prompt dependency highlight areas for improvement.

## Future Work

- Optimize Deep Image Blending for faster performance.
- Explore automated prompt refinement for better segmentation.
- Test on more diverse and complex scenes.

Despite its challenges, the proposed method provides a strong foundation for practical applications in creative design and augmented reality.

# References

[1] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[2] Lingzhi Zhang, Tarmily Wen, and Jianbo Shi. Deep image blending, 2019.

# Acknowledgements