HACETTEPE UNIVERSITY

DEPARTMENT OF ELECTRICS and ELECTRONICS ENGINEERING

ELE492 IMAGE PROCESSING

MIDTERM EXAMINATION REPORT

NAME : **Ziya Tuna**

SURNAME : **Bölükbaşı**

NUMBER : **21628166**

SPRING 2023

Question-4:

Meta AI's "Segment Anything" and Microsoft's Segment-Everything-Everywhere -All-AtOnce" are two significant advancements that have taken place this year. Experiment with their Demo files, run their codes with your own images, and discuss the advantages and limitations of the methods. Additionally, read their papers and try to explain the concepts in your own words.

A) Meta AI’s "Segment Anything" :

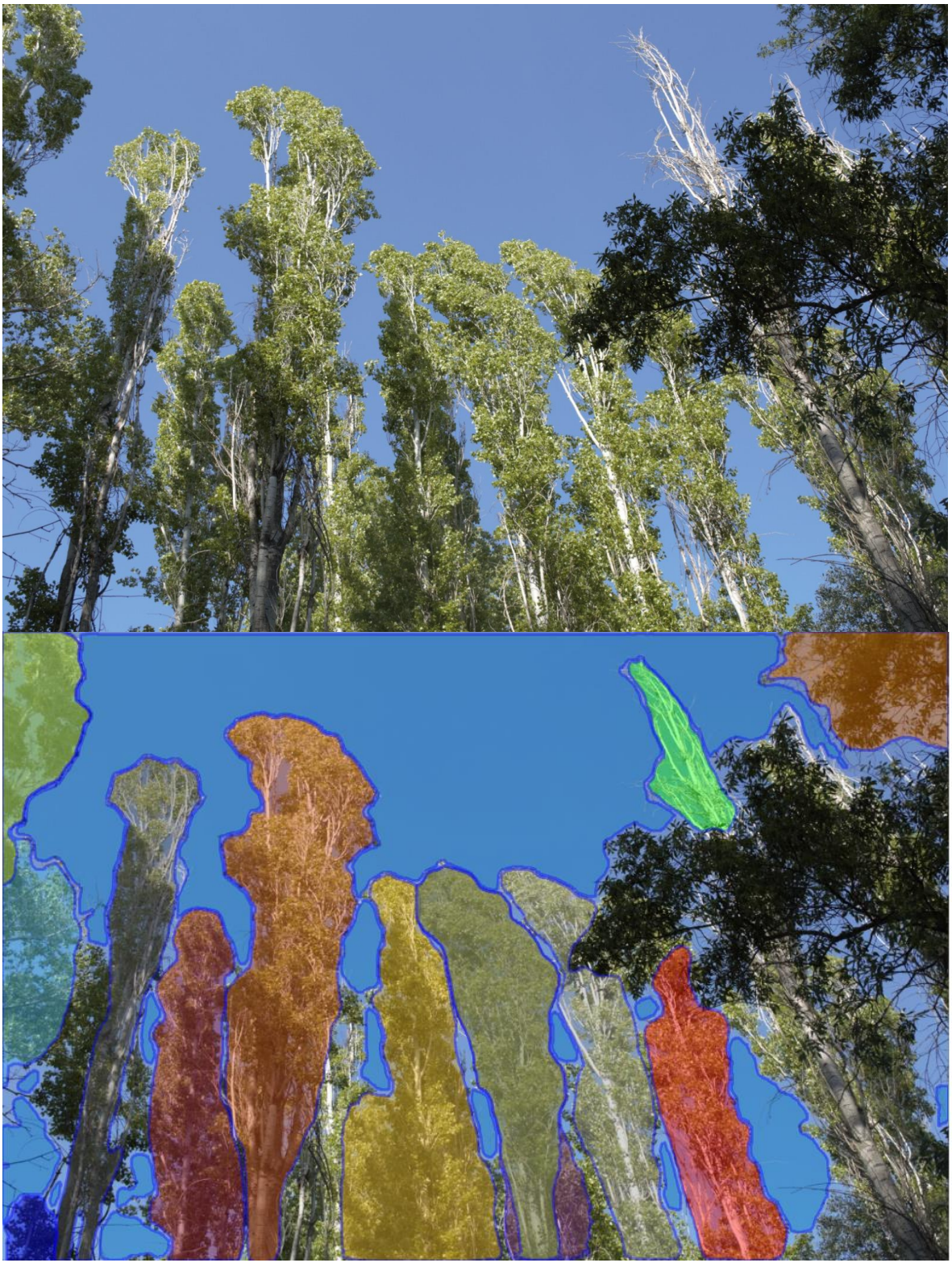

Figure 1: An image I took in the Yeşil Vadi and segmented version of this image by SAM

The SAM model consists of 3 parts: image encoder, prompt encoder and mask decoder. For the image encoder, they use a pre-adapted VisionTransformer. In the prompt encoder, 2 sets of prompts are considered, sparse and dense. For dense coding, convolutions are used. The mask decoder efficiently maps the image embedding, the prompt embedding and the output token to a mask.

The SAM model consists of 11 million different high-resolution images and 1.1 billion high-quality masks.

The SAM model is intended to bring image segmentation into the era of basic models.

Figure 2: A picture of me and its segmented version by SAM

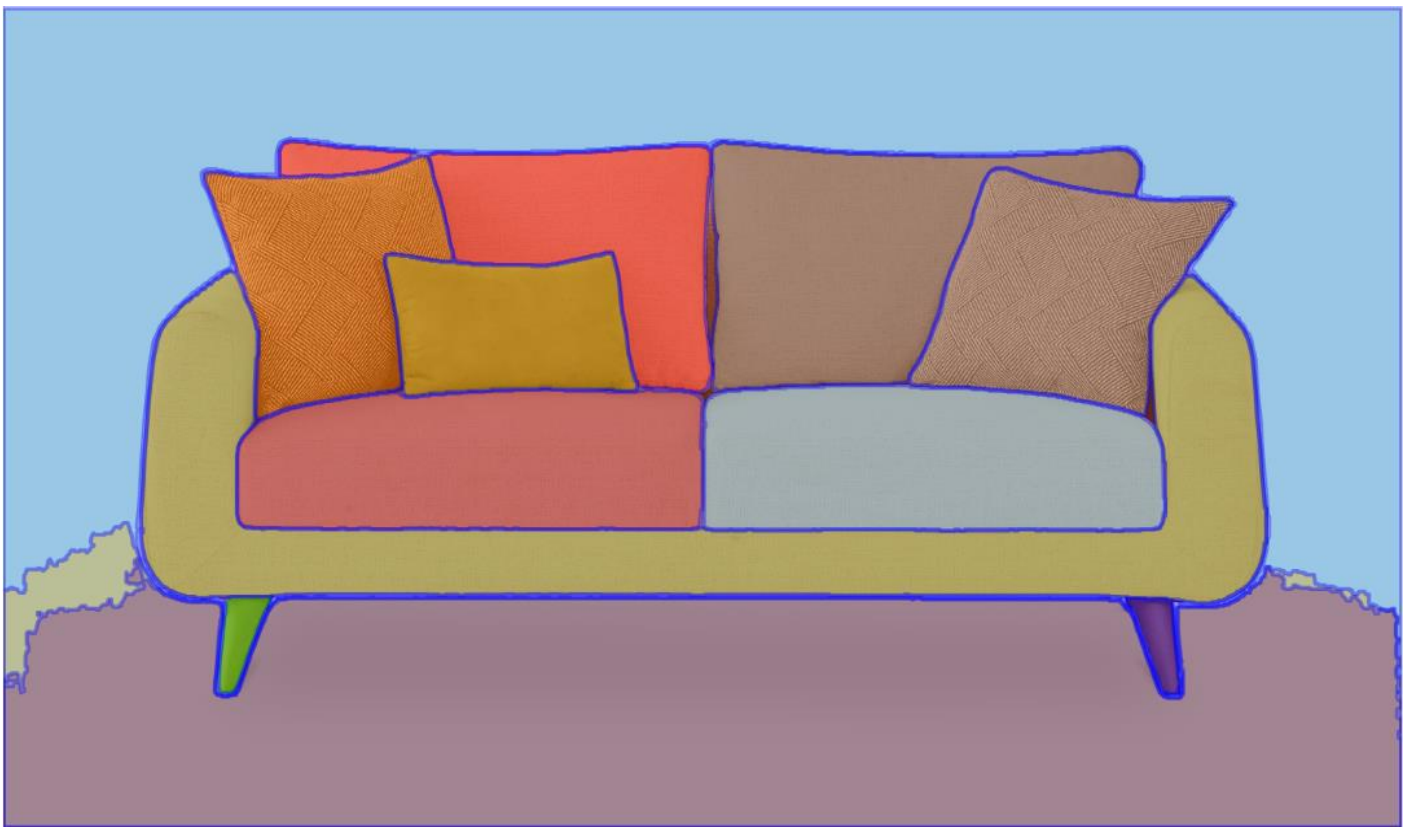

Figure 3 : A seat and segmented version of this seat by SAM

B) Microsoft Segment Everything Everywhere All At Once

SEEM uses a generic encoder-decoder architecture. However, there is a sophisticated interaction between queries and prompts. Given an image, it uses an image encoder to extract image features and predicts masks based on the query output of mask embeddings and class embeddings.

SEEM is trained on small-scale segmentation and cannot perform fragment-based segmentation because it is trained with object-level mask annotations.

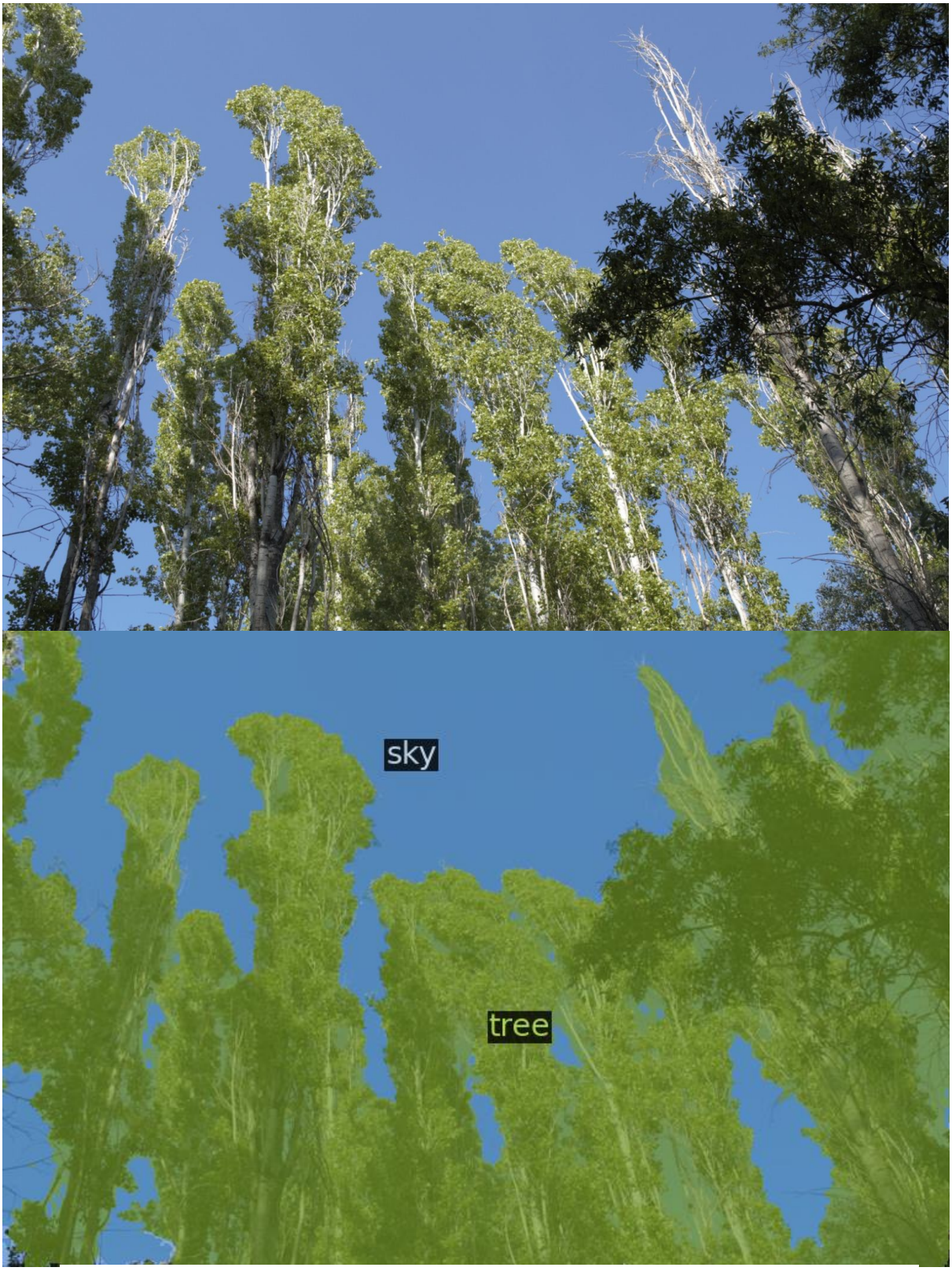

Figure 4: An image I took in the Yeşil Vadi and segmented version of this image by SEEM

Figure 5: A picture of me and its segmented version by SEEM

Figure 6 : A seat and segmented version of this seat by SEEM

C)Comments

Compared to SEEM, I believe SAM segments considerably better. SAM divides images into a lot more pieces. The usage of naming in SEEM, on the other hand, makes it more appealing to the end user.