Istanbul Technical University - Computer Engineering
BLG 561E - Deep Learning
Homework 2
Prepared by: Dr. Yusuf H. Şahin
E-mail: sahinyu@itu.edu.tr
Deadline: 10.05.2025

Question 1 (50 pts) - Learning from Few Data

For many deep learning tasks, using the train set as-it-is is not sufficient for a given method. Thus, the train data can be augmented for a better result. Both pointwise image processing and image transformations can be used as data augmentation for machine learning procedures.

In this part, the data you will use is from ICCV VIPriors Image Classification Challenge[1] which is a subset of ImageNet dataset of 50 different classes. In this challenge, the main focus was using only a small bit of data to learn. **All pretrained models are prohibited!**

Join the Kaggle competition via the following link:

https://www.kaggle.com/t/2f437326c2464afcb6b24b9abbc14d2a

The challenge dataset is a small version of Imagenet which only contains 50 classes. Use the train/test split given in the following Dropbox link:

https://www.dropbox.com/s/856mb0pr5f7e7vl/image-classification.zip?dl=0

Question 2 (50 pts) - Grounded SAM

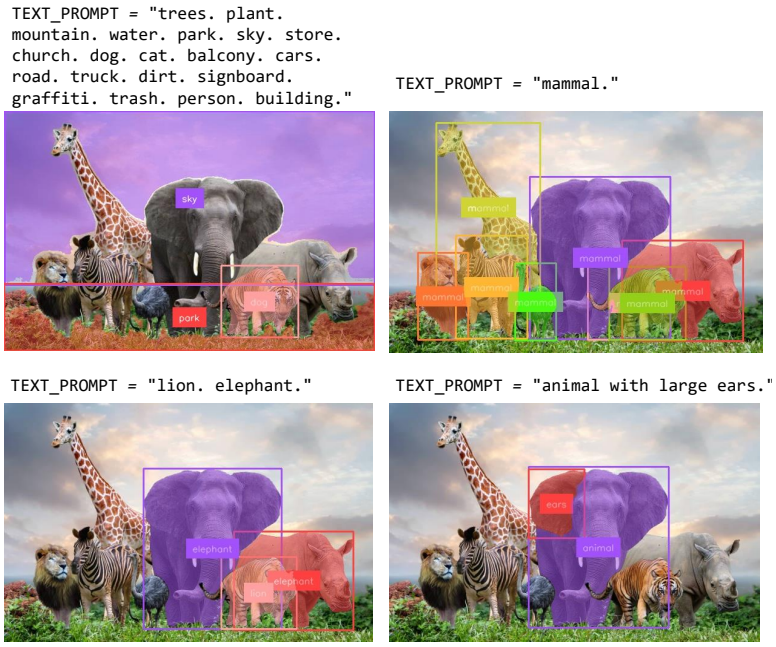

Figure 1: Some segmentations for Grounded SAM for given text prompts.

[1] https://competitions.codalab.org/competitions/33214

Grounded SAM[2] integrates Grounding DINO[3], an open-set object detector, with the Segment Anything Model (SAM)[4] to enable open-vocabulary detection and segmentation. Grounding DINO first identifies objects in an image based on arbitrary text prompts, producing bounding boxes. These boxes are then used by SAM to generate precise segmentation masks. By combining these two models, Grounded SAM can flexibly detect and segment almost any object or region described in natural language.

Use the test dataset from the previous question. Generating different prompts and your own algorithmic logic, classify the examples in the dataset. **Do not train any neural network.**

Join the Kaggle competition via the following link:

https://www.kaggle.com/t/21f309dc3add48ae888fc965c70c0f3d

[2]Ren, T., Liu, S., Zeng, A., Lin, J., Li, K., Cao, H., ... & Zhang, L. (2024). Grounded sam: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159.

[3]Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., ... & Zhang, L. (2024, September). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European Conference on Computer Vision (pp. 38-55). Cham: Springer Nature Switzerland.

[4]Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026).