AI Proposal

Team 4: 17101992 Sumin Hong | 17102041 Dalwon Kim | 20102122 Hyoan Jeong

1. What is the problem that you will be investigating? Why is it interesting? Why do you choose this topic?

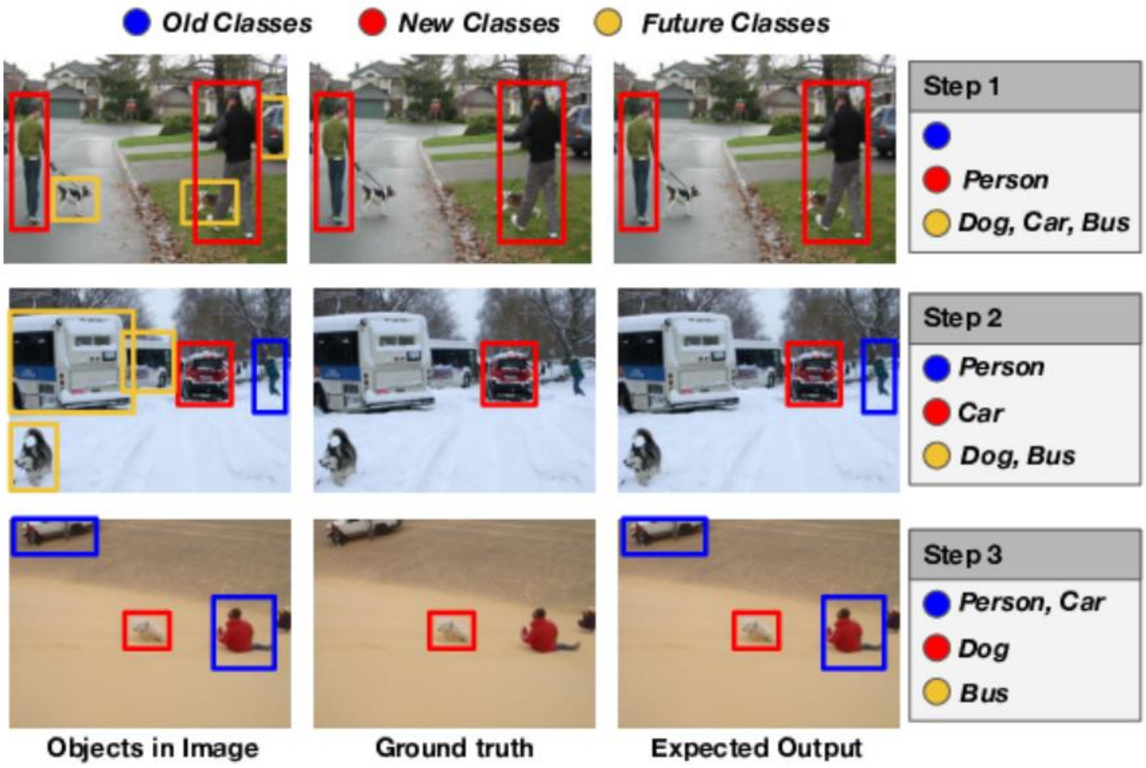

Fig 1. Overall task explanation

We start with the problem that learning on a fixed dataset, typical of deep learning, is not the same as in a real-world setting. If new classes emerge over time that don't exist in the dataset you trained on, the model won't be able to recognize them. The task of solving this problem is called class incremental learning. We can apply this to detailed computer vision tasks, especially object detection, which is a key component of computer vision. This involves identifying and localizing multiple objects within an image using a format called a bounding box. However, applying this to object detection is complicated by the need to identify multiple classes within a single image. We felt that solving this problem is critical in the application of models in the real world, especially in fields that are very close to reality such as robotics, automated driving, and so on, so we were intrigued to choose this topic.

2. What dataset are you using? If needed, how do you plan to collect it? Please describe the dataset as much as you can.

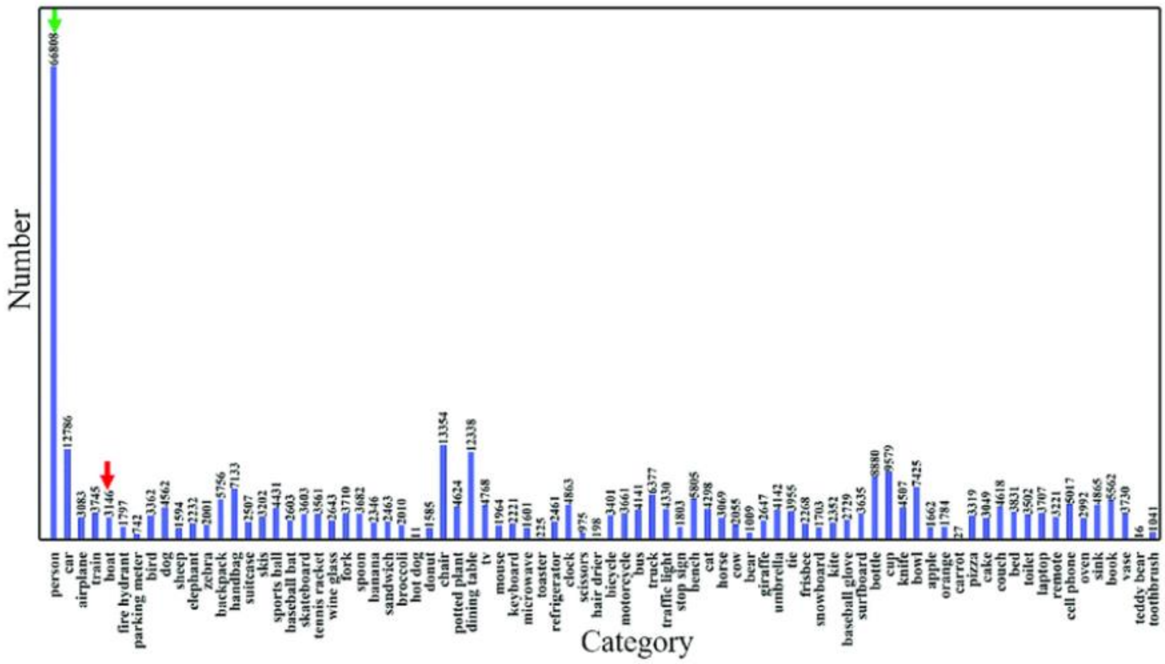

Fig 2. class distribution of the MS COCO 2017

In this work, we will use the MS COCO2017 dataset. The MS COCO 2017 (Microsoft Common Objects in Context) Object Detection Dataset is a large-scale dataset designed for object detection tasks in computer vision. It contains 80 object classes, ranging from everyday items to animals and people. The dataset comprises a large number of images (several hundred thousand), sourced from varied and everyday scenes. And each image in the dataset includes detailed annotations. These annotations consist of object bounding boxes and class labels for each object present in the image. The distribution of images across the 80 classes is uneven, mirroring the natural occurrence frequency of these objects in real-world scenes. This dataset is extensively used for training and benchmarking object detection models, due to the comprehensive nature of its annotations. MS COCO 2017 dataset contains over 200,000 images. This includes both training and validation sets.

3. What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations?

We'll mainly be using replay, which is one of the few methods used by class incremental learning. This is memory replay, where parts of the past dataset are appropriately buffered and used to train the model with new data. The proper organization of this buffer is the main point of a good algorithm. The choice of which past data to store in the buffer varies from paper to paper, mostly using randomized strategies. In contrast, we want to use a strategy based on training loss. Our hypothesis is that data with low training loss will be better remembered later, even if it is trained on less data. We also want to reflect the distribution of classes in the organization of the buffer. As you can see from #2, the class distribution in this dataset is very unbalanced. If we don't properly reflect this, the person class will take up all of our buffer. The previously i mplemented object detection model (there are numerous numbers of object detection model!!) will serve as our baseline. We will build on it to organize our buffers and add the ability to reu se them. Of the object detection models already implemented, we supposed to use the ViT bas ed model and plan to utilize pre-trained weights.

4. **What reading do you examine to provide context and background? What papers (previous works) do you refer to?**

    - Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. iCaRL: Incremental classifier and representation learning. In CVPR, 2017
    - Jeng-Lun Shieh, Qazi Mazhar ul Haq, Muhamad Amirul Haq, Said Karam, Peter Chon dro, De-Qin Gao, and Shanq-Jang Ruan. Continual learning strategy in one-stage obj ect detection framework based on experience replay for autonomous driving vehicle. Sensors, 2020
    - Feng Li, Hao Zhang, Shilong Liu, Jian Guo, Lionel M Ni, and Lei Zhang. Dn-detr: Acc elerate detr training by introducing query denoising. In CVPR, 2022
    - Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. Incremental learning of o bject detectors without catastrophic forgetting. In ICCV, 2017
    - Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. ICLR, 2020
    - Manoj Acharya, Tyler L Hayes, and Christopher Kanan. Rodeo: Replay for online obje ct detection. BMVC, 2020
    - Xialei Liu, Hao Yang, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Multi-task incremental learning for object detection. arXiv, 2022

5. **How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g., pl ots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compar e your results? (e.g., what performance metrics)?**

We use standard COCO evaluation metrics, including the mean average precision (mAP) with Io U threshold, which includes AP, AP.5, AP.75, AP_s, AP_m, and AP_l. This metric is a common ev aluation method for object detection. To utilize this in an incremental setting, we can divide the data into two tasks (past and present), train the model on the current data, and then compare it with no processing at all, and then with the model we want to develop.
For the qualitative evaluation, we can show the predicted images. If we split the data into 2 tas ks (past and present), we can train the model on the current data and use the images that pre dicted the test data. This process will provide us with a quantitative analysis and allow us to co mpare the data without any processing to the data with the model we wanted to develop.