Practical Work 2
Handling open-set-recognition

The objective of this practical work is to examine how image recognition models perform in an open-set recognition scenario—specifically, when encountering both seen and unseen classes. Additionally, the study aims to explore techniques that improve model robustness when predicting unknown objects.

Object detection models are typically trained to recognize a predefined set of known classes with high confidence. However, when these models are faced with unseen objects, they may behave unpredictably. This challenge, known as open-set recognition, has led to several proposed solutions, including:

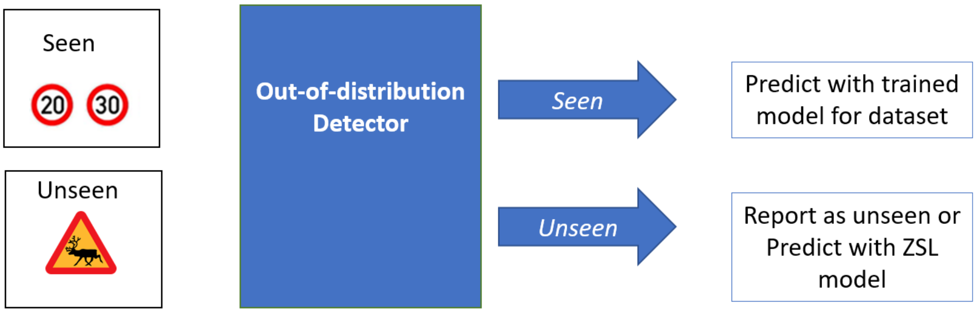
- Out-of-distribution detection: A model designed to determine whether an object belongs to the set of known classes or is an outlier.
- Open-set recognition models**:** These models are trained on both seen and unseen classes, learning to assign a "dummy" label to unknown objects while correctly classifying known ones.
- Zero-shot learning models: Models that specialize in recognizing unseen classes by leveraging word embeddings and image embeddings rather than relying on direct training examples.

The idea of this work is to implement either an out-of-distribution model or an open-set-recognition model as explained in the following (choose ONE option). Please note that the development of the ZSL model is completely optional.

Approach 1 – Out of distribution detection Model

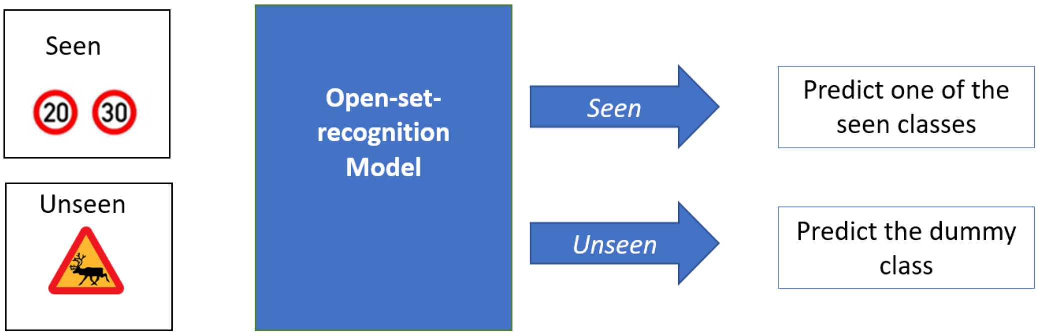
Approach 1 outlines the integration of an out-of-distribution detection model to handle unseen objects effectively:

1) The out-of-distribution detector determines whether the image belongs to the known data distribution.
2) If the image falls within the distribution (i.e., it belongs to a seen class), the trained model predicts its class. Then report the predicted label and confidence level of the prediction.
3) If the image is outside the distribution (i.e., an unseen class): a) Display a warning. b) For informative purposes, report the predicted class and confidence level of the prediction, even though the model was not trained to recognize this class. c) Optionally, predict the unseen class using a zero-shot learning model.



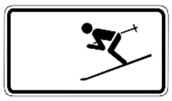
Approach 2 – Open-set-recognition Model

Approach 2 demonstrates how to detect unseen classes using an open-set recognition model. A standard image recognition model trained on a given dataset is retrained to predict a 'dummy' class for unidentified object types. To evaluate its performance, the model should be tested on both known and unknown categories. Alongside the predicted class, the confidence score of the prediction should also be reported.



Example dataset

The German Traffic sign recognition benchmark dataset contains images of 40 types of German traffic signs. Traffic signs from other countries would be unseen classes.

Seen classes	Unseen classes
 **Figure 2.** German traffic sign recognition benchmark (GTSRB) dataset.	  

Description of the work

The experiments should be documented in a report, including the following steps:

1) **Selection of an image data set and object recognition model**
   Select an image data set for the known data distribution. The object recognition model can be either custom-trained or pre-trained, provided that, in the case of a pre-trained model, sufficient information is available regarding the images used for training (training set) and those designated for validation (validation set).

2) **Evaluate the prediction performance** on seen classes within the validation set.
   a) Asses the per-class prediction performance measuring accuracy, precision and recall.
   b) Assess the per-class prediction confidence by analyzing softmax probability scores and entropy measurements.
   c) Describe if there a classes more difficult to recognize for the model.

2) **Evaluate the prediction performance** on unseen classes.
   Analyze the behavior of the image recognition model when predicting instances from unseen classes taking into account the predicted class and confidence level of the prediction.

3) **Approach 1:** Train an **out-of-distribution detection model** and evaluate its performance using a test set containing both instances from seen and unseen classes.
   a) Use an autoencoder based model trained on in-distribution data. Evaluate OOD on the reconstruction error of an image.
   b) Optional: Compare the performance with another OOD approach (any of your choice). For instance [1] provides an alternative approach where that performs OOD measuring the Mahalanobis distance on the latent space. Describe the differences you have observed in the evaluation with the test set, as well as strengths and limitations.
   c) For evaluation, ensure that no images from the training set are used.

4) **Approach 2**: Train an **open-set recognition model** using both known and unknown classes.
   a) Use an adaption of the loss function of the **Negative Log-Likelihood (NLL) Loss** using thresholding or assigning higher penalties for training the prediction of the dummy class (see additional information section) .
   b) For evaluation, ensure that no images from the training set are used.
   c) Evaluate the performance on a test set containing seen and unseen classes, and report the predicted labels along with their confidence scores.
   d) Compare the model’s performance comparing the implementation of losses with different thresholds or penalizations.

5) Summarize your conclusions based on the results obtained. This section should include:
   - Why handling unknown classes in the context of your chosen dataset can be relevant?
   - How was the behavior of the model trained on seen classes for unseen classes?
   - How is the performance of the OOD or open-set-recognition model to detect unseen classes?

   Furthermore, comment on the following points:

   - Positive findings: Highlight aspects that performed well.
   - Negative results: Discuss challenges, limitations, or unexpected behaviors observed in the models.
   - Open issues: Identify areas that require further investigation.
   - Future work: Suggest possible improvements, alternative approaches, or next steps.

Additional notes:

- You do not need to implement the Zero-shot-learning approach (it's completely optional).
- OOD according to [1] is also completely optional. It is an interesting approach, nevertheless the implementation can be challenging.
- We value that the experiments are clearly described.
- Out-of-distribution detection and open-set recognition remain active research areas, so imperfect results are expected and acceptable.
-

Adaption of the negative log-likelihood function:

To enhance the **Negative Log-Likelihood (NLL) function** for handling unseen classes, two possible adaptations are:

Introducing a Confidence Threshold – This ensures that if the model's highest softmax probability **falls below a predefined threshold**, the input is classified as **unknown**, preventing misclassification into a known class.

Applying Higher Penalties for Misclassification – When an **out-of-distribution (OOD) sample** is incorrectly assigned to a known class, the loss function **penalizes the model more severely**, reinforcing better separation between known and unknown categories.

Bibliografic references:

[1] Denouden, Taylor, et al. "Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance." *arXiv preprint arXiv:1812.02765* (2018).