Transfer Learning Project

Chenyu Tian
Columbia University
ct3308@columbia.edu

Yilu Yang
Columbia University
yy3626@columbia.edu

1 Introduction

This project aims to solve a hierarchical classification problem by predicting both a superclass and its corresponding subclass for a given image. The problem has two primary complexities that test model robustness and generalization:

Open Set Recognition (OSR): The test set contains novel superclasses and subclasses not seen during training. The model must identify and reject these “unknown” samples as “novel”.

Distribution Shift: The test set’s class frequencies may differ from the training set. The model must generalize to this new statistical distribution of known classes.

2 Related Work

OSR contrasts with traditional closed-set recognition, which assumes all testing classes are known during training. The methods for solving OSR problems primarily fall into the following two groups [SD23].

Discriminative Models: Discriminative models aim to learn decision rules directly. This is often approached either by learning highly compact and discriminative representations for known classes (through score-based, distance-based or reconstruction-based methods) or by explicitly introducing ”unknown” information into the training process. This unknown information can be synthesized from known classes.

Generative Models: The generative models is another approach which can be further divided into Instance and Non-Instance Generation-based methods [GHC20]. The former method focuses on generating useful new samples, while the second one focuses on learning the underlying distributions of the known-class data, often using Autoencoders (AEs) or Generative Adversarial Networks (GANs). The principle is that unknown samples will not fit the learned distributions, allowing for their detection based on deviation or high reconstruction error.

3 Method / Algorithm

3.1 Dataset Preparation

Training Data with Novel Label

- **Mixup Data Augmentation:** Generate synthetic novel samples by creating linear interpolations between pairs of training images belonging to different known classes.
- **Simulated Open-Set Split:** Simulate an open-set scenario by holding out specific subclasses to serve as novel samples in the validation set.

Distribution Shift

- **Data Augmentation:** Adopt the suggestion from the brief to use data augmentation (e.g., brightness, contrast, minor geometric transforms) to improve generalization.

- **Validation Set:** Build a validation set for hyperparameter tuning.

3.2 Models

Since it is a hierarchical OSR problem, we propose different hierarchical models.

3.2.1 Cascading Probability-Based Threshold Classifier with Conditional Constraints

Problems and proposed solutions that are included in this method:

1. *Do we use different models for superclass classification and subclass classification?*
   Answer: No, we use the same base model (e.g., ResNet) as a feature extractor and train two different heads for superclass classification and subclass classification.

2. *How to determine whether the label is novel or known?*
   Answer: By comparing different threshold methods.
   Threshold Methods:

   (a) Maximum Softmax Probability (MSP) [HG18]: Logit → Softmax → Threshold Gate

   (b) Out-of-Distribution Detector (ODIN) [LLS20]: Logit → Temperature Scaling → Softmax → Threshold Gate

   (c) Energy-based Out-of-distribution Detection [LWOL21]: Logit → Energy → Threshold Gate

   (d) Class-wise Sigmoid with Binary Cross-Entropy (BCE) Loss: Logit → Sigmoid → Threshold Gate

   (e) Auxiliary Confidence Gating [DT18]: Penultimate Features → Confidence Branch → Sigmoid → Threshold Gate

   MSP is the baseline model.
   One problem with MSP and ODIN is that they both rely on the Softmax function. The softmax function forces the output probabilities to sum to 1 and loses the magnitude information of the logit, which leads to the **overconfidence** problem. So threshold methods based on the logit are more robust.
   One problem with most of the threshold methods is that they are not trainable, while Auxiliary Confidence Gating makes it trainable by introducing the second confidence branch besides the main classification branch.

3. *How to let the superclass classification impact the subclass classification?*
   Answer: By applying a soft constraint (Channel-Wise Attention [HSA+19]) during training and a hard constraint (Masking) during inference.

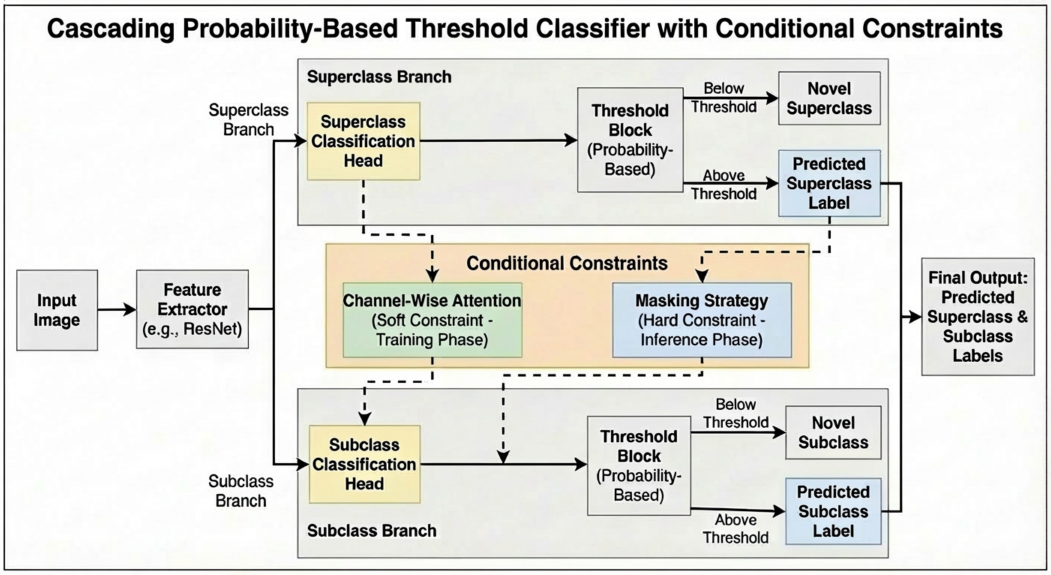
Model Structure: Use a pre-trained feature extractor followed by a superclass classification head and a subclass classification head. Train the network jointly using a combined cross-entropy loss from both heads.

Training and Inference:

- Select a **threshold method** for novel superclass and subclass classification.

- Select a **soft constraint method** used in training, use **Masking** as a **hard constraint method** used in inference.

- Predict the superclass then the subclass based on the threshold method. Apply the constraint method to the subclass classifier to avoid inconsistent labeling.



3.2.2 OpenMax Hierarchical Classifier

As OpenMax [BB16] is a classic discriminative method in OSR, we plan to investigate its adaptation to our hierarchical problem.

- **Stage 1 - OpenMax for Superclass:** Use OpenMax at the superclass level, it will either classify a sample as known superclass or novel superclass.
- **Stage 2 - OpenMax for Subclass:** If the sample is identified as novel at the previous level, it will also be classified as novel subclass. Otherwise, we train a separate OpenMax classifier on its corresponding set of known subclasses.

3.2.3 Hierarchical Class Anchor Clustering (CAC)

- **Stage 1 - Hierarchical Anchor Design:** Design structured anchors [MSMD21] by first defining a unique base vector for each superclass. Subclass anchors will then be created by adding small offset vectors to their corresponding superclass base vector.
- **Stage 2 - Training:** Train the network using CAC loss function with hierarchical anchors forcing the network to learn a semantically meaningful logit space where the cluster layout directly mirrors problem's hierarchy.
- **Stage 3 - Inference:** Use the distance-based rejection process to classify known subclasses, while distinguishing rejected samples as either "Unknown Superclass" or "Unknown Subclass".

3.3 Evaluation

- **Primary Metric:** Our top priority is to significantly improve upon the CLIP baseline's performance on "Unseen Accuracy".
- **Secondary Metrics:** We will monitor all official metrics, including "Overall Accuracy", "Seen Accuracy", and "Categorical Cross-entropy" for both class levels.
- **Threshold Tuning:** To determine the gatekeeper's probability threshold and the expert's confidence threshold, we will simulate an open-set scenario (e.g., by holding out subclasses during training) and tune on our local validation set to maximize "Unseen Accuracy".

References

[BB16] Abhijit Bendale and Terrance E. Boult. Towards Open Set Deep Networks . In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1563–1572, Los Alamitos, CA, USA, June 2016. IEEE Computer Society.

[DT18] Terrance DeVries and Graham W. Taylor. Learning confidence for out-of-distribution detection in neural networks, 2018.

[GHC20] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3614–3631, 2020.

[HG18] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks, 2018.

[HSA+19] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

[LLS20] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks, 2020.

[LWOL21] Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. Energy-based out-of-distribution detection, 2021.

[MSMD21] Dimity Miller, Niko Sunderhauf, Michael Milford, and Feras Dayoub. Class anchor clustering: A loss for distance-based open set recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3570–3578, 2021.

[SD23] Jiayin Sun and Qiulei Dong. A survey on open-set image recognition. *arXiv preprint arXiv:2312.15571*, 2023.