Computer Vision

Multi-label Image Classification

Group 25
Riajul Islam, Andreas Calonius Kreth, Christine Midtgaard

***Abstract*—Here is a summary of the project and the conclusions.**

***Index Terms*—Multi-label learning, deep learning, computer vision, multi-label classification, deep learning for MLC.**

I. Introduction

Multi-label classification is the supervised learning problem where an instance may be associated with multiple labels. Image classification is a computer vision task that requires assigning a label or multiple labels to an image. Single-label classification, or multi-class classification, refers to the problem where an image contains only one object to be identified. However, natural images usually contain multiple objects or concepts, highlighting the importance of multi-label classification [1]. In this project we investigate two methods aimed to solve two different problems within multi-label learning: Query2Label targets the challenges of imbalance and object localization, whereas MLSPL addresses the challenge of training an effective multi-label classifier from minimal supervision.

A. Multi-Label Classification Problem

II. Related Work

Multi-label learning is a well studied problem within computer vision [2].

a) Loss Functions:

b) Convolutional Neural Networks:

c) Vision Transformer:

d) PU learning: Learning from positive and unlabeled data: a survey by Bekker and Davis [3].

Learning to Classify Texts Using Positive and Unlabeled Data by Li and Liu [4].

III. Theoretical Background

The theory behind multi-label classification and the methods in the chosen papers.

A. Query2Label: A Simple Transformer Way to Multi-Label Classification

Query2Label: A Simple Transformer Way to Multi-Label Classification (Query2Label) by Liu et al. [5] is a two-stage framework for multi-label classification. It uses transformer decoders to extract features with multi-head attentions focusing on different parts of an object category and learn label embeddings from data automatically.

a) Backbone architecture: The CvT by Wu et al. [6].

B. Multi-Label Learning from Single Positive Labels

Multi-Label Learning from Single Positive Labels (MLSPL) by Cole et al. [2]. How it uses weak supervision and contrastive learning.

IV. Methodology

Focus Methodology on the rationale for choosing these methods, how the operationalized their training, and any adaptations made for reproduction.

V. Experimental Setup

A description of our setup and the author's setups.

A. Dataset

The MS-COCO 2014 [7] dataset is used as a benchmark for evaluation both Query2Label and MLSPL. MS-COCO (Microsoft Common Objects in Context) is a large-scale dataset commonly used for object detection, segmentation, and multi-label image classification. COCO consists of 82,081 training images and 80 classes, and a validation set of 40,137 images.

B. Implementation Details

1) Query2Label: We evaluated the Query2Label model using the best performing backbone model, the CvT-w24 backbone, with a 384 input resolution, pretrained on the ImageNet-22k dataset, as described by Liu et al. [5]. For our experiments, we used a pretrained model checkpoint released by the authors, which was trained on the MS-COCO 2024 dataset, described in the previous section. The model was tested on the MS-COCO 2024 validation set with a batchsize of 16 and otherwise no fine-tuning or modification.

2) Multi-Label Learning from Single Positive Labels:

a) Data preparation: The dataset preparation for the MS-COCO dataset in MLSPL relies on converting the standard multi-label MS-COCO dataset to a single positive label format, simulating a weakly supervised setting. Cole et al. does this beginning with the fully annotated multi-label image dataset and corrupt it by discarding annotations, simulating single positive training data by randomly selecting one positive label for each training example [2].

To convert into a single positive label format, the instructions provided by Cole et al. are followed: firsly, both images

TABLE I
COMPARISON OF MAP RESULTS BETWEEN OUR IMPLEMENTATION AND REPORTED RESULTS ON THE MS-COCO 2014 DATASET.

Method	Dataset	mAP (Ours)	mAP (Paper)
Q2L	COCO	91.3	91.3
MLSPL	COCO	66.3	66.3

and annotations for training and validation are downloaded. Secondly, pre-extracted features for COCO, provided by the authors are downloaded. Lastly, the script `format_coco.py` is used to produce uniformly formatted image lists and labels.

b) Hyperparameters:

C. Evaluation Metrics

To asses model performance, we adopt mean Average Precision (mAP), a standard metric widely reported in multi-label classification tasks as it is used to analyze the performance object detection and segmentation. Both Query2Label and MLSPL report results in terms of mAP.

VI. RESULTS AND DISCUSSION

This section compares results from the experiments to the those of the papers. Discuss why they might not be the same, or describe the similarities. Discuss whether the methods are able to solve the problems.

A. Query2Label

B. Multi-Label Learning from Single Positive Labels

VII. CONCLUSION

REFERENCES

[1] T. Ridnik, G. Sharir, A. Ben-Cohen, E. Ben-Baruch, and A. Noy, "Ml-decoder: Scalable and versatile classification head," 2021. [Online]. Available: https://arxiv.org/abs/2111.12933

[2] E. Cole, O. M. Aodha, T. Lorieul, P. Perona, D. Morris, and N. Jojic, "Multi-label learning from single positive labels," 2021. [Online]. Available: https://arxiv.org/abs/2106.09708

[3] J. Bekker and J. Davis, "Learning from positive and unlabeled data: a survey," *Machine Learning*, vol. 109, no. 4, p. 719–760, Apr. 2020. [Online]. Available: http://dx.doi.org/10.1007/s10994-020-05877-5

[4] X. Li and B. Liu, "Learning to classify texts using positive and unlabeled data." 01 2003, pp. 587–594.

[5] S. Liu, L. Zhang, X. Yang, H. Su, and J. Zhu, "Query2label: A simple transformer way to multi-label classification," 2021. [Online]. Available: https://arxiv.org/abs/2107.10834

[6] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "Cvt: Introducing convolutions to vision transformers," 2021. [Online]. Available: https://arxiv.org/abs/2103.15808

[7] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015. [Online]. Available: https://arxiv.org/abs/1405.0312