Partial Label Learning meets Active Learning: Enhancing Annotation Efficiency through Binary Questioning

Shivangana Rawat [1] **Chaitanya Devaguptapu** [2] **Vineeth N Balasubramanian** [1]

Abstract

Supervised learning is an effective approach to machine learning, but it can be expensive to acquire labeled data. Active learning (AL) and partial label learning (PLL) are two techniques that can be used to reduce the annotation costs of supervised learning. AL is a strategy for reducing the annotation budget by selecting and labeling the most informative samples, while PLL is a weakly supervised learning approach to learn from partially annotated data by identifying the true hidden label. In this paper, we propose a novel approach that combines AL and PLL techniques to improve annotation efficiency. Our method leverages AL to select informative binary questions and PLL to identify the true label from the set of possible answers. We conduct extensive experiments on various benchmark datasets and show that our method achieves state-of-the-art (SoTA) performance with significantly reduced annotation costs. Our findings suggest that our method is a promising solution for cost-effective annotation in real-world applications.

1. Introduction

Acquiring annotations of superior quality (Deng et al., 2009; Everingham et al., 2010; Cordts et al., 2016) holds utmost importance in the development of machine learning models that excel in achieving state-of-the-art (SoTA) performance across diverse tasks (He et al., 2016; 2017; Chen et al., 2018). Nonetheless, the process of obtaining annotations can be resource-intensive and time-consuming(Settles & Craven, 2008a). Active learning (Settles & Craven, 2008a) emerges as a valuable technique to mitigate the costs and time associated with annotation acquisition by employing an iterative approach of querying a human expert to provide labels for a subset of unlabeled data points (Yoo & Kweon, 2019; Gal et al., 2017; Sinha et al., 2019). The primary objective of active learning is to intelligently select the most informative data points that warrant labeling, thereby facilitating accelerated and accurate learning by the model.

Conventional supervised learning approaches necessitate the assignment of a solitary, accurate label to each training instance (Deng et al., 2009; Everingham et al., 2010; Cordts et al., 2016). Nevertheless, in numerous applications, obtaining a single ground-truth label can prove to be a costly (Settles & Craven, 2008a), time-consuming (Settles & Craven, 2008a), or even unattainable endeavor. An illustrative instance of such a challenge arises when attempting to label images of rare avian species, where the expertise of a qualified ornithologist may be scarce. Moreover, even if experts are accessible, the potential for human error further compounds the labeling process.

In juxtaposition to single labels, the acquisition of partial labels (Cour et al., 2011) proves to be a relatively straightforward task. Partial labels (Jin & Ghahramani, 2002; Nguyen & Caruana, 2008; Liu & Dietterich, 2012; Cour et al., 2011) represent incomplete annotations that lack the comprehensive information required to ascertain the true label of a data point. For instance, consider a partial label assigned to an image of a cat, such as "feline." While this label does not divulge the precise breed of the cat, it does offer adequate information to categorize the data point into the appropriate class. The process of obtaining partial labels is considerably more feasible than obtaining single labels, primarily due to the cost-effectiveness, swiftness, and broader domain expertise possessed by crowdworkers compared to domain-specific experts.

Partial label learning (PLL) (Cour et al., 2011) has gained a lot of traction recently due to its ability to learn from data with partial labels. Most of the existing works focus on learning to disambiguate partial labels and find the true original labels (Jin & Ghahramani, 2002; Nguyen & Caruana, 2008; Liu & Dietterich, 2012; Chen et al., 2014; Yu & Zhang, 2017). However, one of the strong constraints of PLL methods is that the true label has to be a part of the candidate set (Cour et al., 2011). This assumption makes it hard

[1]Indian Institute of Technology, Hyderabad [2]Fujitsu Research, India. Correspondence to: Shivangana Rawat <cs20mtech12001@iith.ac.in>.

AI & HCI Workshop at the 40^{th} International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA. 2023.

to use PLL methods in many real-world applications. However, in this work, we astutely leverage this very constraint as an advantage to ease the annotation process.

In this work, we propose an approach to improve the efficacy of the annotation process by combining active learning and partial label learning. We use active learning to intelligently select the binary questions that are most informative and employ partial label learning techniques to learn from partial labels resulted due to these binary questions. By strategically optimizing the query selection process, our aim is to achieve state-of-the-art (SoTA) performance with minimal questioning. To summarize, our contributions encompass the following key aspects:

1. A Novel Integration: We present a straightforward yet highly effective methodology that combines active learning and partial label learning techniques to significantly diminish the costs associated with annotation. Our approach offers a compelling solution for optimizing the annotation process while ensuring resource efficiency.

2. State-of-the-Art (SoTA) Performance: Our proposed approach not only attains SoTA performance with only a fraction of the entire data, but it also achieves this by asking a substantially reduced number of binary questions compared to approaches that solely rely on either active learning or partial label learning.

2. Related Work

Active Learning: Active learning (Settles & Craven, 2008a) is a machine learning paradigm that can improve the performance of machine learning models by reducing the amount of labeled data required. The model actively queries an oracle for labels on a specific set of data points, this alleviates the cost and time required for annotating the entire dataset while not performing on par or better than models trained with fully-labeled datasets.

To choose what samples to label, various strategies like least confidence (Lewis & Gale, 1994), margin sampling (Scheffer et al., 2001) and entropy sampling (Settles & Craven, 2008b) are used. With the emergence of deep learning, these query strategies were later applied (Wang et al., 2017) on deep neural networks with encouraging results. While new strategies specific to deep neural networks(Yoo & Kweon, 2019; Gal et al., 2017; Sinha et al., 2019) have been proposed, strategies like entropy and random sampling are still strong competitors(Munjal et al., 2020). We refer the reader to (Pengzhen et al., 2020) for a detailed survey of active learning techniques in deep neural networks.

Partial Label Learning: Partial label learning (PLL) (Cour et al., 2011) is a type of weakly supervised learning where each training example is associated with a set of candidate labels, among which only one is the true label. This setting arises in many real-world applications where it is difficult to give one precise label either due to a lack of knowledge of annotation budget or the cost associated with obtaining precise labels might be too high.

Partial label learning methods primarily focus on disambiguating the ground-truth label from the candidate label set associated with each training example. Existing strategies for disambiguation include identification and averaging. Identification based methods (Jin & Ghahramani, 2002; Nguyen & Caruana, 2008; Liu & Dietterich, 2012; Chen et al., 2014; Yu & Zhang, 2017) identify the label that is most frequently associated with the training example. Averaging based methods (Hüllermeier & Beringer, 2005; Cour et al., 2011; Yu & Zhang, 2017) treat all the candidate labels equally and average the modeling outputs for making predictions.

Active Partial Label Learning: While active learning (AL) and partial label learning (PLL) have been extensively investigated separately, there is a scarcity of research that comprehensively combines both methodologies. Existing AL+PLL methods (Zhang et al., 2023; Li et al., 2022) that query labels from an oracle require the oracle to select the label from a large number of classes, which can be difficult. In contrast, our method divides the task of querying partial labels into smaller and easier tasks of answering binary questions. This makes labeling easier for the oracle and also improves the performance of the active learning algorithm. The work by (Hu et al., 2018), focuses on acquiring true labels for all the samples with reduced annotation effort. They train a multiclass classifier from the partial labels in order to take advantage of the partial labels. However, their method does not leverage partial label learning methods for disambiguation purposes. Moreover, the approach assumes access to the class ontology, which is not always available in practical scenarios. Our approach addresses the challenges associated with class ontology availability and disambiguation while combining AL and PLL. We divide the task of querying partial labels into smaller and easier tasks of answering binary questions. This makes labeling easier for the oracle. We also leverage partial label learning methods for disambiguation purposes. This allows us to achieve better performance as compared to baseline approaches.

3. Methodology

3.1. Preliminaries:

Pool-Based Active Learning: Pool-based active learning is the most commonly adopted active learning strategy aimed

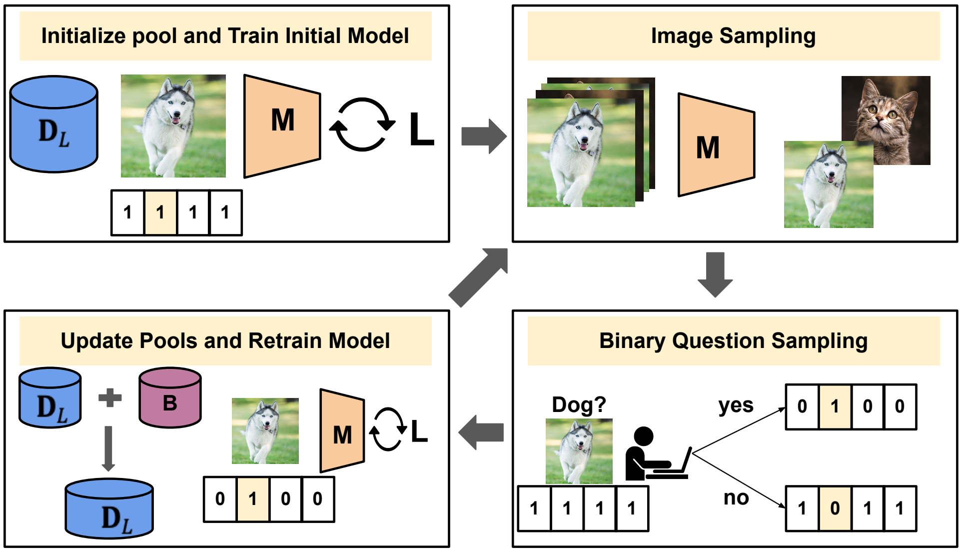

Figure 1. This figure shows the end-to-end pipeline of our proposed strategy. Here M represents the model, D_L represents the labeled pool and B represents the selected batch of samples.

at reducing the number of labeled samples required for effective model training. The process unfolds iteratively, starting with an unlabeled dataset T, which can be partitioned into disjoint sets L and U such that $L \cup U = T$ and $L \cap U = \emptyset$. Initially, set L comprises randomly selected samples from T, which are subsequently labeled by consulting an oracle. The remaining unlabeled samples from T are included in set U.

Model M is trained using the labeled set L. Subsequently, multiple cycles of active learning commence. Within each cycle, a query function Q is employed to sample a subset B from the unlabeled set U. The samples in subset B are then labeled, subsequently removed from the unlabeled set U, and added to the labeled set L. The model M is retrained using the updated labeled set L. This process of active learning cycles is repeated until the model converges or the annotation budget is fully utilized.

Partial Label Learning (PLL) Partial Label Learning (PLL) is a type of weakly supervised learning, where the training set consists of pairs $(x_i, \mathbf{y_i})$, representing the samples and their corresponding candidate label sets. The dataset contains a total of n samples, denoted by $T = \{(x_i, \mathbf{y_i})\}_{i=1}^{n}$, $x_i \in \mathbb{R}^d$ denotes a sample from the training set.

In PLL, labels are defined as sets, which are often referred to as candidate sets $\mathbf{y_i}$,$\mathbf{y_i} \subseteq \{1...C\}$, where C represents the total number of possible classes. The candidate set $\mathbf{y_i}$ captures the potential labels associated with a given sample. Within the PLL framework, it is assumed that each candidate set $\mathbf{y_i}$ includes a true label y_i, expressed as $\mathbf{y_i} = y_i \cup \mathbf{z_i}$, where $\mathbf{z_i}$ represents the distractor set. While we have access to the candidate set $\mathbf{y_i}$, the true label y_i remains hidden.

The primary objective of PLL methods is to train a classifier model M capable of learning the relationship between the input instance x_i and the true label y_i. By leveraging the available candidate sets, The challenge in PLL lies in effectively handling the presence of distractors within the candidate sets and disambiguating the true label from the distractor labels. PLL algorithms aim to infer the underlying true labels accurately.

3.2. Obtaining Partial Labels in an Active Manner

Obtaining Partial Labels through Binary Questions As outlined in Section 1, acquiring a single definitive label can often be costly or even unattainable. To address this, we adopt a strategy of obtaining reliable partial labels through binary questions posed to an oracle. By formulating binary (yes/no) inquiries, we elicit partial labels efficiently.

The process involves presenting the oracle with an (image, class) pair and asking the question, ”Does the specific image belong to the given class?” The oracle provides a binary response of either yes or no. This approach enables us to obtain partial labels in a more resource-conscious manner, alleviating the burden of annotating each sample with a precise class label.

When the binary question receives a ”yes” response, we identify the true label, thereby finding a single label within our candidate set for that image which becomes the true label. Conversely, if the answer to the binary question is ”no,” we can eliminate that particular class from the candidate set associated with the image.

To reduce the overall annotation cost, we can strategically choose binary questions that have a higher likelihood of receiving a "yes" response. By minimizing the number of binary questions required, we enhance the efficiency of the annotation process, optimizing resource allocation.

Pipeline Figure 1 shows the entire pipeline. We are given a set of unlabelled images of size n, a list of C classes which the images must be classified into and a maximum annotation budget of b questions. We divide the annotation budget equally among all the samples and hence each sample gets an annotation budget of b/n questions. Initially, all the images have all the classes in their respective candidate sets. We start by randomly sampling a set of images and adding them to our labeled set of samples. All the remaining samples are added to the unlabeled set U. For every sample in the labeled set L, we ask random binary questions till we have obtained the true label or the per-image budget is exhausted which would provide us with a true label or partial label respectively. We use this partially labeled set L to train a model M. Now we run multiple active learning cycles where we use an image query function Q_i to sample a set of images B. For every image in the sampled set B, we query questions based on a question sampling function Q_q till the true label is acquired or the per-image question budget is exhausted. This partially labeled batch B is added to the labeled set L and removed from the unlabeled set U. The labeled set L is used to retrain a model M. This cycle continues till desired performance is reached.

3.3. Sampling Strategies

3.3.1. Image Sampling Strategies

In this section, we describe the different uncertainty-based active learning approaches.

Random: The random sampling strategy selects samples randomly from the unlabeled pool without considering any specific criterion.

Confidence: The confidence sampling strategy, as introduced in (Lewis & Gale, 1994), selects samples with the least confidence score. The confidence score of a sample is determined by the probability p_m of its highest probability class, calculated as:

$$s = p_m$$

Here p_m is the probability of the highest probability class for the sample.

Entropy: The entropy sampling strategy, proposed in (Settles & Craven, 2008b), chooses samples with the highest entropy score. The entropy score of a sample is computed using the class probabilities p_c, and is defined as:

$$s = -\sum_{c \in \{1 \ldots C\}} p_c \log p_c$$

In the above equation, p_c represents the probability of class c for a given sample.

These sampling strategies enable the active learning process to select informative samples based on their uncertainty, allowing the model to learn from the most challenging and uncertain instances during training.

3.3.2. Question Sampling Strategies

In this section, we discuss different active learning approaches for sampling which class to choose when asking a binary question for annotation. These strategies are used when the true label for a particular sample has not been revealed yet and there is a budget left for that given sample.

Random: The random question sampling strategy selects a class randomly from the set of available classes that have not been asked previously for that sample.

Highest Probability: The highest probability question sampling strategy selects the class with the highest probability according to the current model's predictions. This strategy aims to prioritize the class that is considered most likely by the model.

By employing these question sampling strategies, active learning can effectively determine which class to inquire about, contributing to the iterative learning process and improving annotation efficiency.

Method	Number of Questions	Accuracy
PLL (30%)	134K	79.38 ± 0.39
PLL (60%)	224K	81.50 ± 0.99
PLL (90%)	270K	81.88 ± 0.37
OURS	76K	80.55 ± 0.20

Table 1. The table shows the comparison of our proposed approach with different partial label annotation budgets. Our method outperforms the partial label approaches with respect to the number of questions by a large margin.

4. Experiments:

In this section, we provide details about the experimental setting architectures, the dataset we use to test our approach along with the evaluation criteria.

4.1. Experimental Setting:

Datasets and Training Settings In order to show the effectiveness of our proposed approach, we evaluated it on three standard image classification datasets: MNIST (LeCun et al., 1998), CIFAR10(Krizhevsky et al., 2009) and

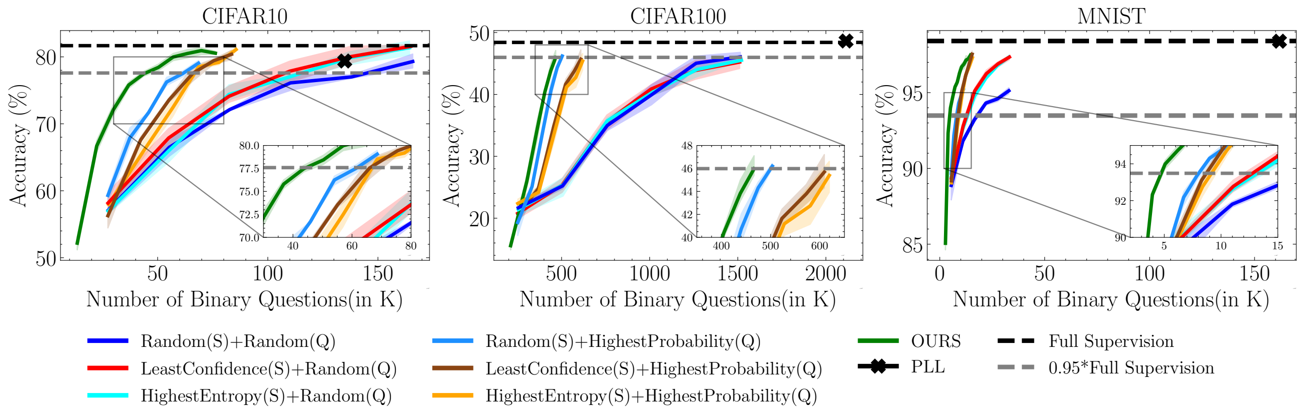

Figure 2. The figure shows the comparison of our proposed approach of partially labeling the dataset actively (OURS) with the vanilla active learning baseline approaches which are shown as Sample selection strategy (S) + Question selection strategy (Q). OURS uses Random as the sample selection strategy and highest probability as the question sampling strategy. We also show a comparison with vanilla partial label learning.

CIFAR100(Krizhevsky et al., 2009). For all MNIST experiments, we train a three-layer MLP (Laine & Aila, 2016) model. For CIFAR10 and CIFAR100, we train a ResNet (He et al., 2016) model with SGD (Robbins & Monro, 1951) optimizer. We use a learning rate of 1e-2 for MNIST and 5e-2 for CIFAR10 and CIFAR100. Next, we provide details about the baselines, evaluation criterion, learning from partial labels, and active learning settings in our experiments.

Baselines: We compare our proposed strategy with traditional active learning approaches which include random, least confidence (Lewis & Gale, 1994) and entropy (Settles & Craven, 2008b) sampling strategies.

Evaluation Criterion: We evaluated the performance of all methods based on the top-1 accuracy of the models and the number of binary questions required to achieve it. This criterion provided a measure of both accuracy and efficiency for the different approaches.

Learning from Partial Labels: To train our model using the partial labels obtained from querying binary questions, we employ the strategy proposed in the PRODEN paper by Lv et al. (Lv et al., 2020). This strategy is designed to iteratively identify the true label and improve the model's performance.

Active Learning Settings: In our active learning experiments, we use a batch size of 1000 images for the MNIST dataset and 5000 images for the CIFAR10 and CIFAR100 datasets. This batch size determines the number of samples selected for annotation in each iteration of the active learning process.

For all our experimental results, we report the average accuracy based on three independent runs. This ensures a more robust evaluation of the performance of our proposed approach across different datasets and random initialization conditions.

4.2. Results

4.2.1. Performance Comparison with Vanilla Active Learning

To demonstrate the effectiveness of our approach, we compare our proposed pipeline with traditional active learning approaches. Figure 2 presents the results of our experiments, and we make the following observations: (1) Our method consistently outperforms the vanilla active learning approaches on all three datasets: MNIST, CIFAR10, and CIFAR100. (2) On the CIFAR100 dataset, we achieve 95% of the accuracy obtained with full supervision, using only 8% of the overall annotation budget. This highlights the efficiency of our approach in achieving high accuracy with limited annotations. (3) For the same top-1 accuracy, our proposed method requires approximately 20% fewer binary questions compared to the vanilla active learning approaches. This reduction in the annotation budget further emphasizes the efficiency of our approach. (4) Although the highest probability question sampling strategy improves the performance of the vanilla active learning approaches, our proposed pipeline still outperforms it. This demonstrates the additional benefits and effectiveness of our strategy beyond simple question sampling. (5) We also outperform the vanilla partial label learning approach by a significant number of questions.

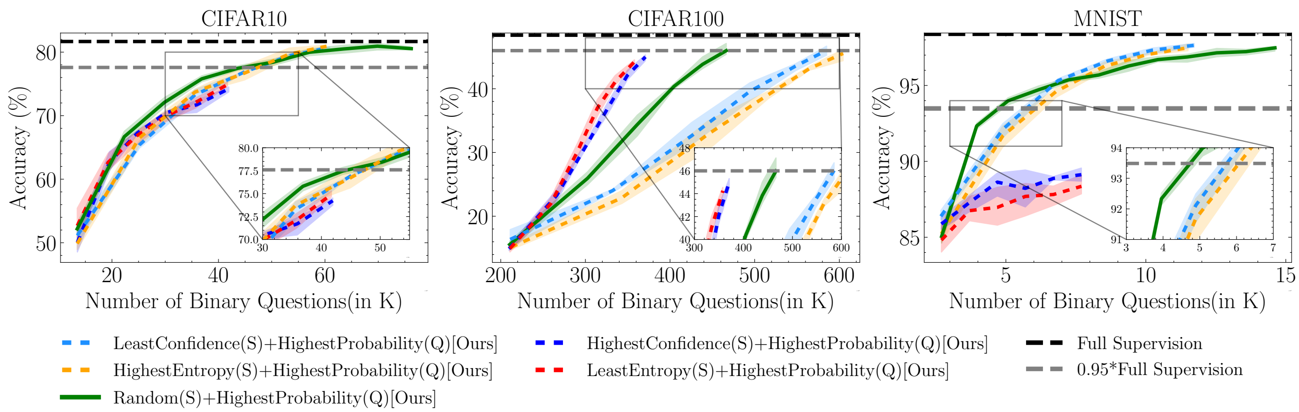

Figure 3. The figure shows the comparison of sample selection strategy on our proposed approach(OURS). The methods are represented as Sample selection strategy (S) + Question selection strategy (Q).

4.2.2. Performance Comparison with Vanilla Partial Label Learning

In Table 1, we compare the performance of our method with vanilla partial label learning approaches on the CIFAR10 dataset. The corresponding results are also indicated by a cross mark in Figure 2 for the other datasets. We observe the following: (1) The performance of the model increases as the number of questions per sample increases from 30% to 60% and then to 90%. (2) Our method achieves similar accuracy but with less than half the annotation budget. This demonstrates the efficiency of our approach in achieving comparable results while significantly reducing the number of binary questions required.

4.2.3. Efficient Image Sampling with Our Proposed Pipeline

We analyze the impact of different image sampling strategies in our proposed pipeline. Figure 3 presents the results of this comparison on all three datasets: MNIST, CIFAR10, and CIFAR100. We compare the image sampling strategies of random, high confidence, low confidence, high entropy, and low entropy. We observe the following trends: (1) Sampling easy samples first provides a performance gain compared to random sampling for CIFAR100. This is because finding the true class in CIFAR100 is a more challenging task, and labeling easy samples initially can help simplify the process. The performance gain obtained from finding the true class outweighs the drop in performance caused by labeling easy samples. (2) However, for MNIST, we observe the opposite trend. The performance gain from finding the true class is unable to compensate for the drop in performance caused by labeling easy samples. Therefore, random sampling performs better than sampling easy samples first in the case of MNIST. These observations highlight the importance of considering dataset characteristics when selecting the image sampling strategy. The effectiveness of different strategies can vary depending on the dataset, and finding the right balance between sampling easy samples and focusing on finding the true class is crucial for optimizing the performance of our proposed pipeline.

5. Conclusion

In this study, we have presented a novel integration of active learning (AL) and partial label learning (PLL) techniques to effectively reduce annotation costs in supervised machine learning. By incorporating binary questions developed through AL+PLL, we streamline the annotation process and obtain high-quality labels while significantly decreasing the overall annotation burden. Through extensive experimentation on diverse benchmark datasets, we have shown that our AL+PLL framework achieves remarkable performance while significantly reducing annotation costs. We believe that our findings pave the way for more efficient and cost-effective annotation strategies that leverage AL and PLL

References

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

Chen, Y.-C., Patel, V. M., Chellappa, R., and Phillips, P. J. Ambiguously labeled learning using dictionaries. *IEEE Transactions on Information Forensics and Security*, 9(12):2076–2088, 2014. doi: 10.1109/TIFS.2014.2359642.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

Cour, T., Sapp, B., and Taskar, B. Learning from partial labels. *Journal of Machine Learning Research*, 12 (42):1501–1536, 2011. URL http://jmlr.org/papers/v12/cour11a.html.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88 (2):303–338, June 2010.

Gal, Y., Islam, R., and Ghahramani, Z. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pp. 1183–1192. PMLR, 2017.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

Hu, P., Lipton, Z. C., Anandkumar, A., and Ramanan, D. Active learning with partial feedback. *arXiv preprint arXiv:1802.07427*, 2018.

Hüllermeier, E. and Beringer, J. Learning from ambiguously labeled examples. pp. 168–179, 01 2005.

Jin, R. and Ghahramani, Z. Learning with multiple labels. In Becker, S., Thrun, S., and Obermayer, K. (eds.), *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002. URL https://proceedings.neurips.cc/paper_files/paper/2002/file/653ac11ca60b3e021a8c609c7198acfc-Paper.pdf.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Laine, S. and Aila, T. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Lewis, D. D. and Gale, W. A. A sequential algorithm for training text classifiers. In *SIGIR '94*, 1994.

Li, Y., Liu, C., Zhao, S., and Hua, Q. Active partial label learning based on adaptive sample selection. *International Journal of Machine Learning and Cybernetics*, 13 (6):1603–1617, 2022.

Liu, L. and Dietterich, T. A conditional multinomial mixture model for superset label learning. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/aaebdb8bb6b0e73f6c3c54a0ab0c6415-Paper.pdf.

Lv, J., Xu, M., Feng, L., Niu, G., Geng, X., and Sugiyama, M. Progressive identification of true labels for partial-label learning. In *international conference on machine learning*, pp. 6500–6510. PMLR, 2020.

Munjal, P., Hayat, N., Hayat, M., Sourati, J., and Khan, S. Towards robust and reproducible active learning using neural networks. *Computer Vision And Pattern Recognition*, 2020. doi: 10.1109/CVPR52688.2022.00032.

Nguyen, N. and Caruana, R. Classification with partial labels. pp. 551–559, 08 2008. doi: 10.1145/1401890.1401958.

Pengzhen, R., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., and Wang, X. A survey of deep active learning. 08 2020.

Robbins, H. and Monro, S. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.

Scheffer, T., Decomain, C., and Wrobel, S. Active hidden markov models for information extraction. In *IDA*, 2001.

Settles, B. and Craven, M. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pp. 1070–1079, USA, 2008a. Association for Computational Linguistics.

Settles, B. and Craven, M. W. An analysis of active learning strategies for sequence labeling tasks. In *EMNLP*, 2008b.

Sinha, S., Ebrahimi, S., and Darrell, T. Variational adversarial active learning. In *Proceedings of the IEEE/CVF*

International Conference on Computer Vision, pp. 5972–5981, 2019.

Wang, K., Zhang, D., Li, Y., Zhang, R., and Lin, L. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(12):2591–2600, 2017. doi: 10.1109/TCSVT.2016.2589879.

Yoo, D. and Kweon, I. S. Learning loss for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 93–102, 2019.

Yu, F. and Zhang, M.-L. Maximum margin partial label learning. *Machine Learning*, 106, 04 2017. doi: 10.1007/s10994-016-5606-4.

Zhang, F., Ye, J., Feng, L., Rao, Z., Zhu, J., Kalander, M., Gong, C., HAO, J., and Han, B. Active learning with partial labels, 2023. URL https://openreview.net/forum?id=1VuBdlNBuR.