AN INTRODUCTION TO DEEP ACTIVE LEARNING (link)

WHAT IS DEEP LEARNING

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. It attempts to build appropriate models by simulating the structure of the human brain.
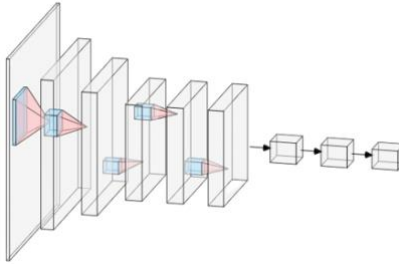

Figure: Structure diagram of convolutional neural network

WHAT IS ACTIVE LEARNING

Active Learning is just a method. It aims to select the most useful samples from the unlabeled dataset and hand it over to the oracle (e.g., human annotator) for labeling, to reduce the cost of labeling as much as possible while still maintaining performance.
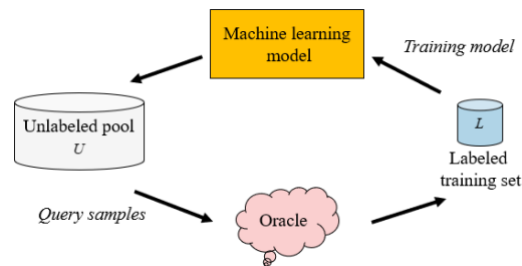

Figure: The pool based active learning

The pool-based AL cycle: Use the query strategy to query the sample in the unlabeled pool U and hand it over to the oracle for labeling, then add the queried sample to the labeled training dataset L and train, and then use the newly learned knowledge for the next round of querying. Repeat this process until the label budget is exhausted or the pre-defined termination conditions are reached.

DEEP ACTIVE LEARNING

Due to the rapid development of internet technology, we have entered an era of information abundance characterized by massive amounts of available data. As a result, DL has attracted significant attention from researchers and has been rapidly developed. However, the acquisition of many high-quality annotated datasets consumes a lot of manpower, making it unfeasible in fields that require high levels of expertise such as speech recognition, information extraction, medical images, etc. Therefore, AL is gradually coming to receive the attention it is due. It is therefore natural to investigate whether AL can be used to reduce the cost of sample annotations. As a result of such investigations, deep active learning (DAL) has emerged.

This combined approach was proposed by considering the complementary advantages of the two methods, and researchers have high expectations for the results of studies in this field.
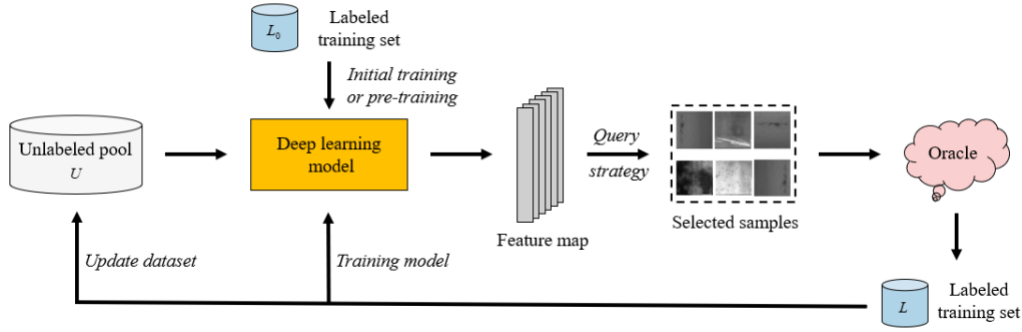
Figure: A typical example of deep active learning

A typical example of DAL: The parameters θ of the DL model are initialized or pre-trained on the label training set L0, and the samples of the unlabeled pool U are used to extract features through the DL model. Then samples are selected based on the corresponding query strategy and the label is queried to form a new label training set L. Later DL model is trained on L, and U is updated at the same time. This process is repeated until the label budget is exhausted or the pre-defined termination conditions are reached.

Though AL-related research into query strategy is quite rich, it is still quite difficult to apply this strategy directly to DL. This is mainly due to:

• *Insufficient data for label samples*: AL often relies on a small amount of labeled sample data to learn and update the model, while DL is often very greedy for data. The labeled training samples provided by the classic AL method thus insufficient to support the training of traditional DL. In addition, the one-by-one sample query method commonly used in AL is also not applicable in the DL context.

• *Model uncertainty*: The query strategy based on uncertainty is an important direction of AL research. In classification tasks, although DL can use the SoftMax layer to obtain the probability distribution on the label, the facts show that they are too confident. The SoftMax response (SR) of the final output is unreliable as a measure of confidence, and the performance of this method will thus be even worse than that of random sampling

• *Processing pipeline inconsistency*: The processing pipelines of AL and DL are inconsistent. Most AL algorithms focus primarily on the training of classifiers, and the various query strategies utilized are largely based on fixed feature representations. In DL, however, feature learning and classifier training are jointly optimized. Only fine-tuning the DL models in the AL framework, or treating them as two separate problems, may thus cause divergent issues.

Approaches to overcome the above difficulties are identified by researchers:

• To address the *Insufficient data* problem, researchers have considered using generative networks for data augmentation or assigning pseudo-labels to high-confidence samples to expand the labeled training set. Some researchers have also used labeled and unlabeled datasets to combine supervised and semi supervised training across AL cycles.

• To solve the *neglect of model uncertainty* in DL, some researchers have applied Bayesian deep learning to deal with the high-dimensional mini- batch samples with fewer queries in the AL context thereby effectively alleviating the problem of the DL model being too confident about the output results.

• To deal with the *pipeline inconsistency* problem, researchers have considered modifying the combined framework of AL and DL to make the proposed DAL model as general as possible, an approach that can be extended to various application fields. This is of great significance to the promotion of DAL.

Query Strategy Optimization in DAL

• *Batch Mode DAL (BMDAL):* The main difference between DAL and classic AL is that DAL uses batch-based sample querying. In traditional AL, most algorithms use a one-by-one query method, which leads to frequent training of the learning model but little change in the training

• *Uncertainty-based and hybrid query strategies*: This approach is simple in form and has low computational complexity and hence a very popular query strategy in AL. It is mainly used in certain shallow models like SVM or KNN. This is mainly because the uncertainty of these models can be accurately obtained by traditional uncertainty sampling methods like Margin Sampling, Least Confidence and Entropy.

• *Deep Bayesian Active Learning (DBAL):* DBAL combines Bayesian convolutional neural networks (BCNNs) with AL methods to adapt BALD to the deep learning environment, thereby developing a new AL framework for high-dimensional data. This approach performs Gaussian prior modeling on the weights of a CNN, and then uses variational inference to obtain the posterior distribution of network prediction.

• *Density-based Methods:* The term, density-based method, mainly refers to the selection of samples from the perspective of the set (core set). The construction of the core set is a representative query strategy. This idea is mainly inspired by the compression idea of the coreset dataset, and attempts to use the core set to represent the distribution of the feature space of the entire original dataset, thereby reducing the labeling cost of AL.

COMMON FRAMEWORK DAL

Cost-Effective Active Learning (CEAL) is one of the first works to combine AL and DL to solve the problem of depth image classification. CEAL merges deep convolutional neural networks into AL, and consequently proposes a novel DAL framework. It sends samples from the unlabeled dataset to the CNN step by step, after which the CNN classifier outputs two types of samples: a small number of uncertain samples, and many samples with high prediction confidence. A small number of the uncertain samples are labeled by the oracle, and the CNN classifier is used to automatically assign pseudo-labels to many high-prediction-confidence samples. Then, these two types of samples are used to fine-tune the CNN and the update process is repeated.
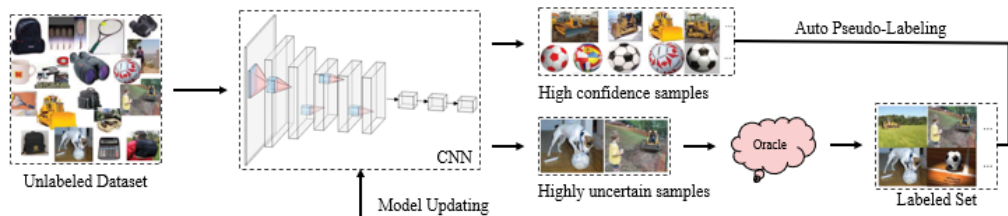

Figure: Overall framework of CEAL

A Deep Active Learning Framework for Biomedical Image Segmentation proposes a framework that uses a fully convolutional network (FCN) and AL to solve the medical image segmentation problem using a small number of annotations. It first trains FCN on a small number of labeled datasets, then extracts the features of the unlabeled datasets through FCN, using these features to estimate the uncertainty and similarity of unlabeled samples. This strategy helps to select highly uncertain and diverse samples to be added to the labeled dataset to start the next stage of training.

Active Palmprint Recognition proposes a similar DAL framework as that for the palmprint recognition task. The difference is that, inspired by domain adaptation, Active Palmprint Recognition regards AL as a binary classification task: it is expected that the labeled and unlabeled sample sets have the same data distribution, making the two difficult to distinguish. Supervision training can be performed directly on a small number of labeled datasets, which reduces the burden associated with labeling.

Learning Loss for Active Learning (LLAL) designed a small parameter module of the loss prediction module to attach to the target network, using the output of multiple hidden layers of the target network as the input of the loss prediction module. The loss prediction module is learned to predict the target loss of the unlabeled dataset, while the top-K strategy is used to select the query samples. LLAL achieves task-agnostic AL framework design at a small parameter cost, and further achieves competitive performance on a variety of mainstream visual tasks (namely, image classification, target detection, and human pose estimation)
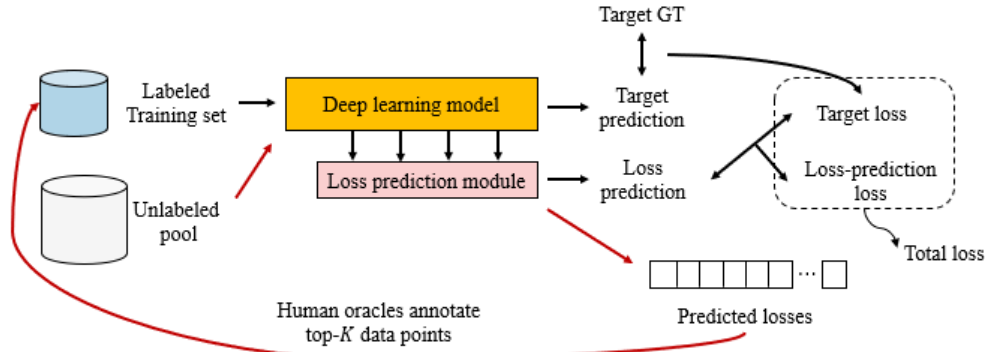

Figure: The overall framework of LLAL. The black line represents the stage of training model parameters, optimizing the overall loss composed of target loss and loss-prediction loss. The red line represents the sample query phase of AL. The output of the multiple hidden layers of the DL model is used as the input of the loss prediction module, while the top-K unlabeled data points are selected according to the predicted losses and assigned labels by the oracle.

APPLICATIONS OF DAL *- A systematic and detailed overview of existing DAL-related work from an application perspective.*

Visual Data Processing: the first field in which DAL is expected to reach its potential is that of computer vision
- Image classification and recognition
- Object detection and semantic segmentation
- Video processing

Natural Language Processing (NLP):
- Sentiment Analysis
- Question-answering and summarization

The other applications include, but are not limited to, gene expression, robotics, wearable device data analysis, social networking, ECG signal analysis, etc.

DISCUSSION AND FUTURE DIRECTIONS

DAL combines the common advantages of DL and AL: it inherits not only DL's ability to process high-dimensional image data and conduct automatic feature extraction, but also AL's potential to effectively reduce annotation costs. DAL therefore has fascinating potential especially in areas where labels require high levels of expertise and are difficult to obtain. The current research directions regarding DAL methods focus primarily on the improvement of AL selection strategies, the optimization of training methods and the improvement of task- independent models. The improvement of AL selection strategy is currently centered around considering the query strategy based on uncertainty and diversity in an explicit or implicit manner. Moreover, hybrid selection strategies are increasingly favored by researchers. Task independence is also an important research direction, as it helps to make DAL models more directly and widely extensible to other tasks. However, the related research remains insufficient, and the corresponding DAL methods tend to focus only on the uncertainty-based selection method. In general, DAL research has significant practical application value in terms of both labeling costs and application scenarios; however, DAL research remains in its infancy at present, and there is still a long way to go in the future.

REFERENCES

[1] A Survey of Deep Active Learning by Pengzhen Ren, Yun Xiao , Xiaojun Chang, Po-Yao Huang, Zhihui Li, https://arxiv.org/pdf/2009.00236.pdf

[2] Nabiha Asghar, Pascal Poupart, Xin Jiang, and Hang Li. 2016. Deep Active Learning for Dialogue Generation. arXiv: Computation and Language (2016).

[3] Hamed H Aghdam, Abel Gonzalez-Garcia, Joost van de Weijer, and Antonio M López. 2019. Active learning for deep detection neural networks. In Proceedings of the IEEE International Conference on Computer Vision. 3672–3680.

[4] Ahmed Hussein, Mohamed Medhat Gaber, and Eyad Elyan. 2016. Deep Active Learning for Autonomous Navigation. (2016), 3–17

[5] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. 2019. Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. arXiv: Learning (2019).

[6] Erik Bochinski, Ghassen Bacha, Volker Eiselein, Tim J W Walles, Jens C Nejstgaard, and Thomas Sikora. 2018. Deep Active Learning for In Situ Plankton Classification.

[7] Anfeng Cheng, Chuan Zhou, Hong Yang, Jia Wu, Lei Li, Jianlong Tan, and Li Guo. 2019. Deep Active Learning for Anchor User Prediction. (2019), 2151–2157.