Active Learning

1.ESA: Annotation-Efficient Active Learning for Semantic Segmentation

Problem Statement: Semantic segmentation has always required assigning class labels to each pixel in an image, which takes exhaustive resources and a high cost of annotation. Current active learning methods attempt to save annotation effort by selecting either informative pixels or small regions for labeling. However, these methods still have lot of defects in that they neglect the larger patterns and structural information inherent in the images.

They involve as many as 5000 annotation clicks for a single image, even when only a small portion of the image for instance, 2% of it is to be selected for labeling.

How does it solve the problem?

The solution introduced by the paper is **Entity-Superpixel Annotation (ESA)**, The research introduces a new active learning strategy to make the annotation process more effective. ESA uses a class-agnostic mask proposal network with superpixel grouping to extract structural patterns in an image. This method prioritizes the superpixels that show high entropy to represent uncertainty, so annotators can identify the most critical entities within each image. In this way, ESA optimizes the number of clicks. This significantly reduces the cost of annotation while keeping the representation comprehensive with respect to the important regions of the image.

The ESA combines SA based on superpixel-based annotation with the selectivity of entity-based annotation. The ESA approach selects, for every image in a target domain, a subset of the entities to which the detailed annotation shall be provided. All superpixels in these selected regions will be labeled fully for completeness, while the approach of EA completes this by selecting only a few key entities for efficiency.

Contributions and Novelty

Even the previous semantic segmentation methods solved the problem of annotating by active learning, but even the small image area of 2% requires 5000 queries to the image. It is a very time-consuming process. The proposed method reduces the number of clicks required for image annotation by **98%** compared to traditional approaches, making it much more efficient. The performance of the segmentation models also improves by **1.71%**, demonstrating that the method not only reduces workload but also enhances the quality of the results. Furthermore, ESA's annotator-friendly design, which groups pixels into superpixels and focuses on key entities, is a step forward from existing pixel-based methods that fail to consider the structure within images.

Downsides

This method could struggle with objects that are small or thin, which may not be captured effectively by superpixel-based techniques. This could limit the precision of segmentation in cases where fine-grained details are crucial.

2.Enhancing Text Classification through LLM-Driven Active Learning and Human Annotation

Problem Addressed by the Paper:

It addresses the cost associated with the annotation of data towards text classification, especially in most Natural Language Processing applications. Large datasets are hard to annotate and require many costs.
It represents the first-ever attempt to combine human base evaluators with LLM and Active Learning within a single framework. This is the very first comprehensive evaluation of the utility and efficiency of combining human annotators, Active Learning, and GPT 3.5 utilities in a text classification task. In this paper, we improve traditional Active Learning methodologies by introducing uncertainty measurements from an LLM such as GPT 3.5 into our annotation selection process. This will help in reducing the cost of manual annotation by taking the strong points of a machine learning model for efficient and effective text classification.

Solution Proposed:

The authors propose a method for incorporating human annotators and GPT 3.5 to enhance the efficiency of text classification tasks. This framework works by the selective choice between human and machine-generated annotations, in a way based on uncertainty measurements drawn from GPT 3.5. In the case of an Active Learning process, uncertainty sampling is included that may involve selecting the most ambiguous data points and subsequently either annotating them with GPT 3.5 or human expertise. This method balances the trade-off between cost efficiency and classification accuracy. They evaluated performance on three public benchmark datasets: IMDB for sentiment analysis, a Fake News dataset for authenticity discernment, and a Movie Genres dataset for multi-label classification. That has shown that LLMs will reduce annotation costs while sustaining or improving accuracy.

Contributions and novelty
The key thing in the paper is the creating a hybrid active learning framework that leverages both human effort and LLM's. It combines the strengths of both of them. The novel integration allows the system to switch dynamically between human and machine annotations based on the

level of uncertainty. It opens a path toward a framework that effectively combines the strengths of both in reaching high performance while reducing costs.

One of the key contributions in their work was the creation of a proxy-validation set. This subset of the total data was used for the estimation of model performance in each iteration of the Active Learning process as a set of labeled samples. It also simulates the statistical distribution presented in the main unlabeled pool and has gone through updates with each iteration. Model accuracy was computed on the proxy-validation set at each iteration of Active Learning. To make this consistent, they took the same percentage of confidence for low-confidence data removal as had been done on the main unlabeled pool and applied it on the proxy-validation set. The remaining data in the proxy-validation set gave us an estimate of the main pool's accuracy-a vital measure when true labels of the pool were unavailable.

Downsides

The main downside of this paper i think is reliability on GPT 3.5. LLM’s doesnt work on some datasets and also LLM’s often lie sometimes. The accuracy of these confidence scores can vary depending on the dataset and task, which may lead to incorrect annotations, particularly when GPT-3.5's confidence is low. The study found that for data points where GPT 3.5's confidence fell below a certain threshold, the rate of incorrect annotations increased substantially.

The fact that considerably more tokens are used for larger datasets and that the cost of the GPT 3.5 API cannot be ignored. Additionally, the framework's reliance on pre-trained models like GPT-3.5 may not be easily adaptable to specific domains where highly specialized annotations are required.

3.Distribution Discrepancy and Feature Heterogeneity for Active 3D Object Detection

The paper discuss solution for the challenge of high annotation cost for LiDAR-based 3D object detection, which is one of the enabling technologies in self-driving and robotics. Annotating LiDAR has multiple object identifications in three-dimensional space and is time-consuming and expensive. That limits the efficiency with which models will learn from data. It proposes an Active Learning scheme, Distribution Discrepancy and Feature Heterogeneity, or DDFH, which reduces annotation costs by selecting the most informative samples concerning geometric features and model embeddings.
Unlike a simple 2D image where you just mark an object with a box, annotating in 3D means you are working in a space where objects have depth, volume, and varying distances from the sensor. This makes the annotation process not only painstakingly slow but incredibly expensive.
The current approaches simply can’t scale without exploding costs. Even though several strategies like automatic labeling or domain adaptation have been attempted to reduce these

costs, they come with their own problems. Automatic labeling often leads to inaccuracies, and domain adaptation struggles when models are applied to new or different environments.

Solution Proposed

The method of DDFH introduced in this paper provided a solution for the problem related to high annotation costs for 3D object detection, focused on two major aspects: **Instance-Level Distribution Discrepancy (DD)** and **Frame-Level Feature Heterogeneity (FH)**. These mechanisms assess the novelty and informative value of samples to allow the model to focus on the most useful data for annotation. Geometric features of objects are combined with model embeddings, developing a more complete understanding of the data. The framework also provides a **Quantile Transform** for normalizing different informativeness indicators, allowing consistency in the choice of the best samples across various scales.

Contributions and Novelty

The major contribution of the paper is the introduction of the DDFH method that considerably raises the cost-efficiency of LiDAR-based 3D object detection. This work has shown that DDFH reduces the annotation cost by **56.3%** compared to state-of-the-art methods, with improved detection performance of **1.8%**. Another new contribution is the mechanism of **Confidence Balance** that makes sure the classes with imbalanced data get proper attention in the course of annotation. The framework further aggregates several indicators, geometrical features, and model embeddings using a **Quantile Transform** in order to handle the issue of how to aggregate features with different scales.

Downsides of the Approach

One shortcoming of the approach is the dependence of this approach on specific hyperparameters for example, the number of components in the Gaussian Mixture Model and the perplexity for t-SNE which may have to be tuned quite carefully regarding a particular dataset. Another important point is that while the method is quite robust across types of models and datasets, the improvement may depend upon the complexity of the environment one aims to model. Finally, this paper focuses a lot on LiDAR-based object detection; this network may find lesser general applicability in domains where 3D object detection is not of high importance. While the improvement using this method will be much better concerning costs for annotation, it may still not scale well when there are extremely large datasets in an environment.