HOMEWORK 1 - CSE 584

Mega Sri Shyam Bollini
PSU ID - 991264286
mpb6512@psu.edu

Paper 1: Scalable Active Learning for Object Detection

i) Problem:

This paper addressed the challenge of labeling large datasets in the context of autonomous driving. The authors aim to build a scalable production system for active learning, specifically focusing on object detection. The paper highlights the difficulty of obtaining high-quality labeled data for training Deep Neural Networks (DNNs) and proposes an active learning framework that optimizes the process by selecting the most informative images for labeling, reducing the overall amount of labeled data needed while maintaining or improving model performance.

The motivation is driven by the need to improve data efficiency for supervised learning methods in autonomous driving systems, where vast amounts of data are collected, but only a subset can be labeled due to time and resource constraints. The proposed system seeks to automate the data selection process, enhance performance with fewer labeled images, and scale the solution for production use, particularly in object detection applications.

ii) Solution:

The paper solves the problem of efficiently labeling large datasets for object detection in autonomous driving systems by implementing a scalable active learning framework;

- *Active Learning Loop:* The core idea of active learning is that the model selects the most informative images from a pool of unlabeled data to be labeled. This reduces the need to label the entire dataset while still achieving high model performance. The system increases the training set in each iteration by actively selecting data that maximizes learning potential, minimizing the overall labeling effort.

- *Scoring Functions:* The model evaluates the informativeness of unlabeled images using several scoring functions. These include entropy, mutual information, and gradient-based methods, which help quantify the uncertainty in the model's predictions. The higher the uncertainty, the more valuable the image is considered for labeling.

- *Sampling Strategies:* To avoid labeling similar images (which may not add much value), the system uses diverse sampling strategies such as k-means++, Core-set, and Sparse Modeling. These strategies ensure that the selected batch of images is diverse, preventing redundancy and making the labeling process more efficient.

- *Scalability:* The framework is designed to handle datasets of significant size, with experiments involving over 2 million images. This makes it suitable for production environments where large-scale data is collected continuously, such as autonomous driving.

- *Iterative Training and Evaluation:* In each iteration of the active learning loop, the model is retrained with the newly labeled data, and its performance is evaluated. This iterative process helps the model improve incrementally by focusing on the most challenging data, rather than a random selection.

- *Production Use Case:* The system was tested in a real-world scenario where the objective was to improve the detection of vulnerable road users (like pedestrians) under challenging conditions, such as low-light environments. By using the active learning system, the selected data led to a significant improvement in detection performance, proving its practical application.

iii) Novelty:

- The paper presents a production-ready, *scalable active learning* system specifically designed for object detection in autonomous driving. Most prior work focuses on classification tasks, but this system tackles the more complex task of object detection, which involves both classification and localization.

- The paper introduces and evaluates *multiple scoring functions* such as entropy, mutual information (MI), gradient-based uncertainty, and bounding box confidence for selecting the most informative images for labeling. These methods help quantify uncertainty and prioritize data selection more effectively.

- The system incorporates *diverse sampling strategies* like k-means++, Core-set, and Sparse Modeling to ensure that the labeled dataset is varied and not redundant. This novelty helps avoid over-sampling similar images, which is especially useful in sequential data like video frames from autonomous driving.

- Unlike many existing studies, which focus on relatively small datasets (e.g., 80K images), this paper applies the active learning framework to *large-scale datasets* with over 2 million images. It demonstrates the ability to handle and process large amounts of real-world, unlabeled data.

- The paper contributes by showing how active learning can be performed *iteratively*, where the model is *retrained* at each step with newly labeled data, allowing for continuous improvement without needing to label the entire dataset at once.

- The system *outperforms manual data selection* by human experts, demonstrating that the active learning framework is not only more efficient but also more effective in identifying challenging images for model training.

iv) Limitations:

- The system assumes high-quality annotations for the labeled data. However, in real-world scenarios, the presence of noisy or incorrect labels could negatively impact the model's performance, which is not addressed in the paper.

- Some scoring functions, particularly those using bounding boxes (Det-Ent), tend to bias the selection towards images containing many objects. This can increase the labeling cost and might lead to an overrepresentation of object-dense scenes in the training set, potentially skewing the model's performance.

Paper 2: Introducing Geometry in Active Learning for Image Segmentation

i) Problem:

The paper addresses challenges faced in obtaining labeled training data for image segmentation tasks, especially in specialized fields such as biomedical imaging. Traditional methods for segmentation require significant manual annotation, which is time-consuming and often requires experts, whose time is scarce and expensive. The paper aims to streamline the annotation process by exploiting geometric priors in 3D image volumes and developing an active learning (AL) approach that selects the most informative voxels to be labeled, minimizing human intervention while maximizing the training efficiency.

ii) Solution:

The authors propose a novel Active Learning framework that integrates geometric priors into the annotation process, specifically tailored for 3D image segmentation. By leveraging these geometric constraints, the approach selects image regions that are easier and faster for experts to label, ensuring that voxels lie on well-defined 2D planar patches. This reduces the complexity of navigating through the 3D volume and allows experts to annotate with minimal effort.

The framework introduces two main contributions:

- A method to exploit geometric priors to select the most informative voxels for labeling.
- A mechanism that simplifies the annotation process in 3D image volumes by allowing users to annotate entire 2D planar patches with only a few clicks. This makes annotating 3D volumes no more difficult than annotating standard 2D images.

The authors validate their approach on various datasets, including EM and MRI volumes, demonstrating a significant increase in annotation efficiency and segmentation performance compared to several baseline methods.

iii) Novelty:

The novelty of this paper lies in its combination of traditional Active Learning uncertainty measures with a new geometric-based uncertainty measure, designed to exploit the spatial coherence of neighboring voxels. Specifically:

The approach introduces Geometric Uncertainty, which considers the likelihood of neighboring supervoxels sharing the same label.

- This reduces the chance of selecting isolated, difficult-to-label voxels and instead prioritizes regions where smoothness in labels is more likely.
- A batch-mode query selection process is introduced, where the system selects and presents a planar patch of voxels for annotation rather than individual random voxels scattered across the 3D volume. This strategy improves both the annotation efficiency and the model's learning rate.

This method outperforms traditional AL techniques by making it easier for the annotator to provide feedback while also leveraging the geometric properties of the data for better segmentation results.

iv) Limitations:

- The method relies on selecting 2D planar patches, which may not always be ideal for all types of 3D data, especially where object boundaries are complex or do not conform to planar structures. In such cases, the geometric constraints might limit the flexibility of the annotation process.
- Although the authors optimize the selection process with a branch-and-bound approach, the computational cost of selecting optimal patches may become prohibitive as the size of the 3D volume increases. Further optimization might be required to ensure real-time performance for larger or higher-resolution datasets.
- The approach is demonstrated on specific datasets (EM and MRI), and its generalization to other types of image volumes, or more complex multi-class segmentation tasks, remains to be explored in future work.

Paper 3: [Deep Bayesian Active Learning with Image Data]

i) Problem:

Active learning (AL) is a powerful technique for reducing the amount of labeled data required for machine learning models. However, traditional AL methods face significant challenges when applied to high-dimensional image data. Deep learning models, which excel in handling large datasets, are typically not used in active learning because they rely on vast amounts of labeled data and struggle to represent model uncertainty, which is essential for AL. These limitations have hindered the application of deep learning-based active learning in areas where acquiring labeled data is expensive and time-consuming, such as medical imaging.

ii) Solution:

This paper introduces Deep Bayesian Active Learning for high-dimensional image data by integrating Bayesian deep learning with active learning. Specifically, the authors develop a framework using Bayesian Convolutional Neural Networks (BCNNs), which can handle small datasets while accurately representing uncertainty. The key idea is to use Bayesian inference techniques like dropout to approximate uncertainty in model predictions, allowing for more informed decision-making on which data points should be labeled next. This approach reduces the number of labels required to achieve high accuracy, making it feasible to apply AL to large-scale image tasks like MNIST digit classification and skin cancer diagnosis.

iii) Novelty:

The novelty of this paper lies in:

- *Integration of Bayesian Deep Learning:* The use of Bayesian CNNs for active learning is innovative because it allows the model to capture both epistemic (model) and aleatoric (data) uncertainty, which are crucial for selecting the most informative data points in AL.
- *MC Dropout for Uncertainty:* The paper leverages MC Dropout, a method that approximates Bayesian inference by performing dropout at both training and test time. This enables the model to estimate its uncertainty and make better AL decisions.
- *Acquisition Functions:* Several acquisition functions, including BALD (Bayesian Active Learning by Disagreement), are used to select the most informative images for labeling, based on uncertainty estimates. This is a departure from traditional kernel-based AL methods and enables the framework to be scalable to high-dimensional image data.

iv) Limitations:

- Performing dropout at test time for Monte Carlo sampling adds computational overhead, particularly when training large CNNs. This makes the approach resource-intensive, especially for large datasets like those encountered in real-world applications (e.g., medical imaging).

- Although the framework is designed for small datasets, the experiments on real-world applications such as melanoma diagnosis are conducted on relatively small datasets. It remains unclear how the method would perform on larger and more diverse datasets where labeling noise and class imbalance are more pronounced.
- The authors reset the model after each acquisition step to avoid local optima, which adds to the training time. While this helps isolate the effects of acquisition functions, it may not be practical in settings where real-time or continuous learning is required.