Active Learning Literature Survey[1]

1. What problem does this paper try to solve, i.e., its motivation
   In this paper, the authors aim to explore how active learning techniques can help to minimize the number of informative labeled data for machine learning training where the machine learning algorithms don't have to label all instances randomly or passively.
2. How does it solve the problem?
   This paper provides a comprehensive survey of active learning techniques, especially for different query strategies that choose the most informative data points for labeling which include uncertainty sampling, query-by-committee, expected model change, variance reduction, and density-weighted methods.
3. A list of novelties/contributions
   Comprehensive Literature Survey:
   - Active Learning Scenarios
   - Query Strategies
   - Theoretical and Empirical Analysis
   - Problem Setting Variants
4. What do you think are the downsides of the work?
   While the paper provides theoretical analysis, it doesn't cover enough practical implementations for the methods.

From Theories to Queries: Active Learning in Practice[2]

1. What problem does this paper try to solve, i.e., its motivation
   The paper surveys the practical challenges of deploying active learning systems in real-world scenarios. While traditional research has shown that active learning can reduce the number of labeled examples needed for machine learning, the transition from theoretical to practical use has revealed new problems such as noisy oracles, variable labeling costs, querying in batches, and other issues not fully addressed in earlier work.
2. How does it solve the problem?
   The paper discusses the practical challenges in six directions:
   - Batch-mode active learning: Instead of querying one instance at a time, queries are made in batches, balancing informativeness and diversity.
   - Handling noisy oracles: The strategies to account for oracle errors and noise, including querying multiple annotators to denoise training data.
   - Cost-sensitive learning: Address the issue of varying labeling costs, proposing methods for predicting and managing these costs to minimize overall expenditure.

- Alternative query types: New types of queries are explored, such as feature-based queries, which may reduce annotation costs.
- Multi-task learning: Querying a single instance for multiple learning tasks to improve efficiency.

3. A list of novelties/contributions
   The paper provides a discussion of the six practical challenges of active learning: querying in batches, noisy oracles, variable labeling costs, alternative query types, multi-task learning, and changing model classes.
4. What do you think are the downsides of the work?
   Some of the solutions, particularly those related to batch-mode learning and multi-task learning, may introduce significant complexity and computational overhead

Deep bayesian active learning with image data [3]

1. What problem does this paper try to solve, i.e., its motivation
   The paper addresses the challenge of applying active learning to deep learning models, especially with high-dimensional data like images. Traditional active learning methods rely on small datasets and model uncertainty for data selection, but deep learning typically depends on large amounts of labeled data and often struggles to represent uncertainty. The goal of this paper is to integrate recent advances in Bayesian deep learning into an active learning framework to improve data efficiency and handle high-dimensional image data effectively.
2. How does it solve the problem?
   The paper proposes the method using Bayesian convolutional neural networks (BCNNs) for active learning, which can represent model uncertainty. The authors design an acquisition function that allows the system to choose informative data points for labeling. This approach reduces the need for large labeled datasets, and makes it suitable for image data.
3. A list of novelties/contributions

- Bayesian CNNs for Active Learning
- Uncertainty-based Acquisition Functions
- Real-world Application

4. What do you think are the downsides of the work?
   The method involves resetting the model and retraining it multiple times during the acquisition process, leading to long training times (20 hours for melanoma experiments).

References

[1] B. Settles, "Active learning literature survey," 2009.

[2] B. Settles, "From theories to queries: Active learning in practice," in *Active learning and experimental design workshop in conjunction with AISTATS 2010*, 2011: JMLR Workshop and Conference Proceedings, pp. 1-18.

[3] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *International conference on machine learning*, 2017: PMLR, pp. 1183-1192.