

EFFICIENT DATA CURATION USING ACTIVE LEARNING FOR A VIDEO-BASED FIRE DETECTION

Keyur Joshi

Ruhr University Bochum, Germany & Bosch Sicherheitssysteme GmbH, Germany

Angelina Aziz

Ruhr University Bochum, Germany

Philip Dietrich

Bosch Sicherheitssysteme GmbH, Germany

Prof. Markus König

Ruhr University Bochum, Germany

ABSTRACT: Video-based fire detection is a crucial object detection problem that relies on accurate and reliable data to detect fires. However, collecting and labeling fire-related data can be time-consuming and expensive, making it difficult to obtain sufficient data for training machine learning models. To address this challenge, uncertainty-based active learning techniques can be used to iteratively select the most informative samples for labeling. This can reduce the amount of labeled data needed to achieve high model performance and has the potential to even prune the training data with fewer informative samples. The traditional sampling-based uncertainty estimation methods are computationally expensive. Hence, an efficient prior network-based ensemble distillation State-of-the-Art approach is evaluated on an internal dataset which still requires relatively higher overhead computation making it difficult for production deployment. A biased softmax differencing-based uncertainty approach and a feature-based hard data mining approach are proposed and compared with the distillation approach. The novel approaches are found to have a very low overhead uncertainty estimation time compared to the ensemble distillation approach and traditional sampling techniques. The methods are evaluated in the context of curating the unlabeled pool data and improving the training data. For completeness, the experiments are performed on three different data sizes, and overall, the frame-wise selection strategy is proved to be better than the sequence-wise querying strategy. The Principal Component Analysis (PCA)-based hard data mining outperformed other methods and improved the model performance by 16.33% with $AUC_{2\%}$ metric when compared with the random selection of data. The approach even outperformed the main network trained on full data by 7.33%, henceforth improving the training data by using informative 26.39% data. The results indicate that novel data mining provides efficient training and pool data curation.

KEYWORDS: Uncertainty Estimation, Active Learning, Object Detection, Outlier Detection, Feature-based cluster analysis, Video-based Fire Detection

1. INTRODUCTION

Traditional smoke detectors require a volume of smoke to reach the detector location and hence generally have a high detection time. Thermal cameras (Sousa, et al. 2020) on the contrary are quite expensive and work in the infrared spectrum resulting in fire detection only when there is significant heat produced. There is no visual confirmation of the fire when using thermal cameras. Hence, deep learning-based video detection can be the solution to decrease the detection time and detect fires based on patterns in the video rather than heat produced. However, for an industrial Deep Learning (DL) application of Fire Detection (FD), the speed and the reliability of the predictions are of major concern. The non-reliability can lead to high economic losses and even human endangerment. Late detection can cause heavy economic and even human losses. Speaking about statistics in the industrial setting in the USA, 1.2 Billion \$ in economic loss along with 16 deaths and 273 injuries occur annually (Campbell, 2018) indicating the importance of reliable video-based fire detection. However, in order to have a reliable DL model for detecting fire, the data selected for the model training should be of high quality.

The research takes inspiration from Peter Norvig's quote: "More data beats clever algorithms, but better data beats more data". The traditional thinking of improving a DL model by increasing the quantity of the data in object detection does not answer why the new data is being added to the training data. Nevertheless, even if large amounts of data is readily available, in object detection problem, the requirement of labeled data raises the cost of annotation massively. In order to improve the performance of the model and simultaneously decrease the cost of labeling, the most informative data have to be sampled.

The aim of decreasing the annotation costs can be achieved by answering two questions: which data has to be selected and why? The passive learning models which receive the data randomly or by humans do not consider the informativeness associated with the data. However, the information is associated with uncertainty related to the data. With the help of uncertain information linked with the data, using Active Learning (AL), the informative data can be sampled iteratively. Sampling informative data using AL requires the estimation of uncertainty. The uncertainty can be referred to as a negative reliability score while a higher uncertainty score means low reliability. The uncertainty can help increase the reliability of the DL models by sampling informative uncertain data from the large dataset, thereby using AL to decrease the cost of the annotation. The concept of AL can assist in the data selection process or autonomously select data, which can subsequently be reviewed and labeled by a human annotator for subsequent training sessions.

The major contributions of our work can be summarized as follows:

- We propose two different methods for estimating uncertainty. One of the methods focuses on estimating uncertainty using feature space and the other approach uses softmax differences for uncertainty estimation.
- We compare different methods based on the time required for uncertainty estimation and the performance when implemented in an iterative AL setting.
- We experimentally show that our novel approach is efficient w.r.t time for uncertainty estimation. The approach even outperforms the State-of-the-Art (SOTA) approach on the AUC metric in an active learning setting.
- We evaluate the performance of different methods w.r.t improving the full training data set by decreasing the data using AL.

2. RELATED WORK

Uncertainty estimation can be used as an additional component for the DL model to increase the trust and robustness of the SOTA architectures. DL models are often black box models, with often limited or no interpretability of the results. With the predictions of the DL model, uncertainty estimates can be incorporated to increase the reliability of the results from the black box models. Gal and Ghahramani (Gal & Ghahramani, 2016) introduced a Monte Carlo Dropout-based (MCD) uncertainty quantification method that has a very low overhead computation cost. The drop block-based uncertainty estimation (Deepshikha, *et al.* 2021) was proposed using Monte-Carlo DropBlock (MCDB). Deep ensembles (DeepEns) (Lakshminarayanan, *et al.* 2017) was used to estimate the predictive uncertainty using model re-training. The Test-Time-data Augmentation (TTA) (Manivannan, 2020) was compared with MCD, MCDB, and DeepEns in the context of uncertainty estimation. The research in the domain of uncertainty quantification is usually done in the fields of model and data uncertainty separately. However, one of the first methods combining the effects of both epistemic and aleatoric uncertainty was proposed in (Kendall & Gal, 2017). The loss function to estimate both uncertainties was suggested for the depth regression and segmentation tasks. Malinin and Gales proposed to estimate the predictive uncertainty of the deep learning model using Prior Networks (Malinin & Gales, 2018) which incorporated explicit modeling of distributional uncertainty. This approach parameterized the Dirichlet network over the categorical distribution to maintain the distribution extraction capability from the student model after the knowledge distillation (Malinin, *et al.* 2019). The Bosch internal research method of a low overhead FACER (Schorn & Gauerhof, 2020) based prior network (FacerDir) uncertainty quantification method will be used in the thesis as one of the SOTA methods for benchmarking. However, due to the high uncertainty estimation computation time, we propose novel approaches with both lower estimation time and more reliable uncertainty estimates.

With the need for better-performing object detection methods, the requirement of the data for training the SOTA architectures has also increased. In order to prioritize the labeling of the data and data curation, a lot of research has been done in the field of Active Learning. A comparison of various acquisition functions for computing uncertainty was done in the setup of AL (Nguyen, *et al.* 2022). Choi *et al.* (Choi, *et al.* 2021) proposed a probabilistic approach to estimate both the model and data uncertainty in a single pass and later performed Active Learning. The scalability and transferability were tested over the probabilistic Active Learning approach. An adaptive framework for active learning was developed and proposed in (Desai, *et al.* 2019) which performed adaptive switching between strong and weak supervision.

The present research in the fields of uncertainty estimation and AL will serve as the inspiration for the proposed paper, which aims to incorporate various techniques into the internal architecture of the Video-based Smoke Detection (VSD) model.

3. METHODOLOGY

An overview of the AL framework using different uncertainty estimation strategies using a business chart is depicted in Fig. 1. The chart illustrates the brief methodology of the approaches in an AL setting.

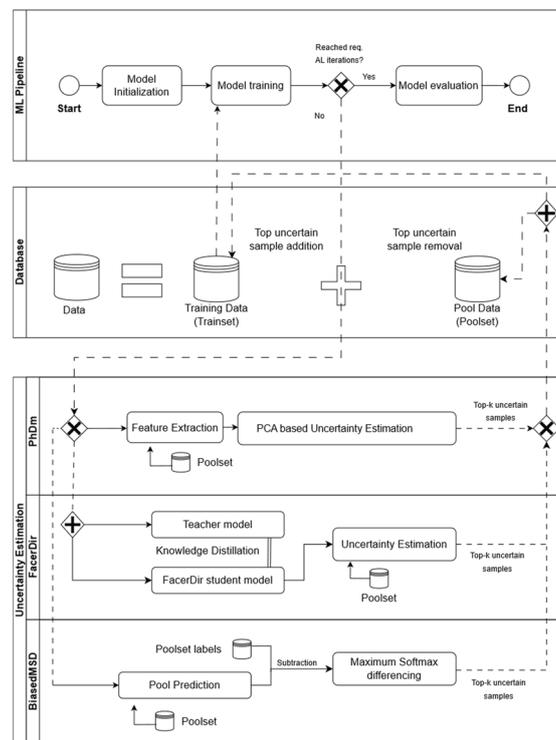


Fig. 1 Thesis methodology which describes the active learning framework implementation using various uncertainty estimation techniques.

3.1 Dataset

The dataset used in the thesis was developed by the engineering team at Bosch. As this dataset is a Bosch internal dataset, it is not available to the public. The dataset comprises Smoke and Non-smoke, “Negative” videos.

The dataset includes video sequences shot over 100 different locations. The locations can be classified into three major scenarios, viz. indoor, outdoor, and semi-outdoor. The scenarios which are shot in an indoor setting ranging from a parking lot to an industry are considered Indoor scenario. The outdoor scenario is a scenario shot in an open environment. The semi-outdoor scenario is on the contrary a setting in which there is a ceiling but lacks enclosure from all sides, resulting in a partially open space. Wind may be present in the outdoor setting and this changes the behavior of the smoke significantly.

3.2 Uncertainty Estimation

3.2.1 PCA-based Hard Data Mining (PhDm)

This approach is a novel method that is based on Principal Component Analysis (PCA) (Jolliffe & Cadima, 2016) for detecting hard samples from the pool data. PCA is a dimensionality reduction approach used to project the data to a low-dimension space and provide interpretability to the data. The PCA is applied to the feature vector space and reduced to the dimension of two. For every image sample, only the feature vector of the maximum prediction is used for dimensionality reduction as shown in Fig. 2.

The outliers found in the PCA plot were investigated and they were found to be often either hard or out-of-distribution examples. Hence, a density-based outlier detection method has been implemented to extract outliers from the PCA plot. Local Outlier Factor (LOF) (Breunig, *et al.* 2000) was used for detecting outliers in the plot which looks at the local density for each point. To query top outlier samples from the plot, a distance-based ranking is performed which ranks each outlier point based on the distance from the nearest inlier point.

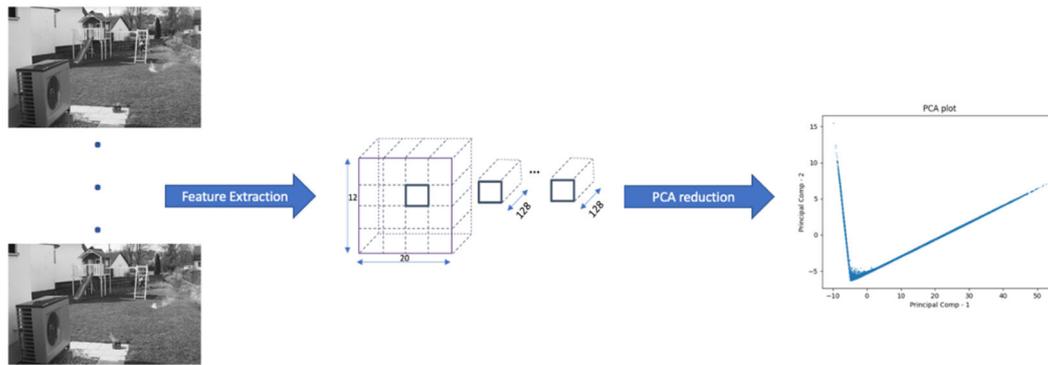


Fig. 2 Feature vector dimensionality reduction. The features (middle) are extracted through the model for the input images (left) and later PCA-based dimensionality reduction is performed. The bold patch represents the patch with maximum softmax prediction whose feature vector is used for dimensionality reduction.

3.2.2 Biased Maximum Softmax Difference (BiasedMSD)

Typically, the final layer of a deep learning classification model is the softmax layer, which produces predictions for a given input. Ideally, the model should accurately replicate the ground truth of the input sample. However, in practice, an entirely perfect or ideal model is unattainable, and a prediction close to the expected ground truth is typically observed.

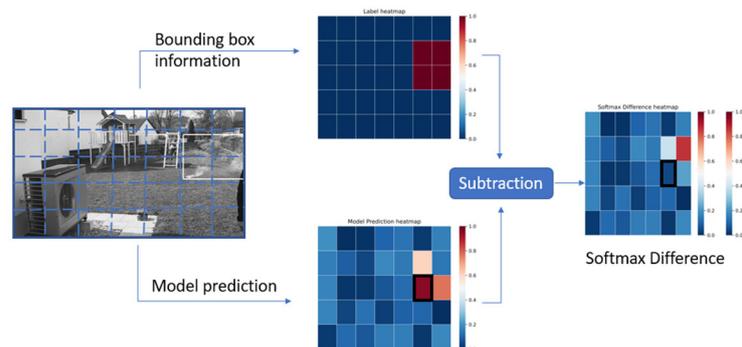


Fig. 3 Illustration of BiasedMSD approach. The difference between the model prediction (mid-bottom) and the ground truths (mid-top) is performed and termed as softmax difference (Bold patch is the patch with maximum prediction and it's softmax difference is used for the whole frame).

The Biased Maximum Softmax Difference is another novel approach that is quite simple and easy to implement. The definition of uncertainty that is adopted in this approach is the distance between the ground truth and reality. For the patch-wise classification problem, as the prediction obtained is patch-wise, the softmax difference is also obtained patch-wise for individual frames.

The drawback of the approach is that one needs the labels beforehand to estimate uncertainty. Due to this shortcoming and the requirements of the labels beforehand, the method is called **Biased Maximum Softmax Difference (BiasedMSD)**, while the bias of requiring ground truth is present for estimating the uncertainty. The visual representation of the working of the BiasedMSD is illustrated in Fig. 3. The inherent bias to require labels for uncertainty estimation suggests that the method may only be used for curating labeled data.

3.3 Active Learning

The final step of the methodology in the scenario of Active Learning (AL) encompasses the implementation of all previously described efficient methods. AL initiates with random sequences in the training dataset, while the remaining data is stored in the pool set. The central concept of AL involves adding a budget (k) of samples or videos to the training set iteratively and removing samples from the pool set. As a result, the iterative process leads to an increase in the size of the training set and a decrease in the size of the pool set. The training of the model is performed on a set of training data, and the methods for estimating uncertainty are utilized to compute scores of uncertainty. The top- k selection for video-based data can be performed using two distinct approaches.

3.3.1 Sequence-wise selection

The sequence-wise (SW) selection method involves selecting the whole video or sequence. Since uncertainty estimation is performed on individual frames, a sampling strategy is employed to aggregate the frame-wise uncertainty estimates into sequence-wise estimates. These sequence-wise uncertainty estimates are then ranked, and the top k sequences are selected. However, this approach may be influenced by a limited number of uncertain frame samples within a video, leading to a bias towards those samples.

3.3.2 Frame-wise selection

The frame-wise selection (FW) method directly ranks the frame-wise uncertainty estimates, and the top- k frame samples are selected. This approach eliminates the need for a sampling strategy, as the frame-wise uncertainty estimates are directly considered.

After the top- k selection, the samples or videos selected from the pool set are added to the training set and removed from the pool set as illustrated in Fig. 1. This whole process is repeated in an iterative manner.

4. EXPERIMENTATION

We conduct different experiments to evaluate and compare the performance of various uncertainty estimation methods in the context of Active Learning. Inception-v1 (Szegedy, *et al.* 2014) was used as a model backbone architecture and the input video was resized into the shape of (640, 360) grayscale images. We use Adam optimizer while training the model.

4.1 Evaluation Metric

In general, object detection algorithms are evaluated using mean Average Precision (mAP) which requires the computation of Intersection over Union (IoU) over the bounding boxes. But the model implemented in the research does not provide bounding box information as an output of the model, but rather a patch-wise classification. Hence, the Receiver Operating Characteristic (ROC) (Streiner & Cairney, 2007) curve has been adopted for performing the model evaluation. ROC curve is computed by evaluating the predictions of the model over different thresholds, and plotting True Positive Rate (TPR) and False Positive Rate (FPR) for each threshold in a curve. As in realistic applications, the FPR should be very low in order to avoid a large number of false alarms. Hence, the Area Under the ROC Curve value under the threshold of 2% FPR ($AUC_{2\%}$) is evaluated and used as a metric to evaluate different methods.

4.2 Uncertainty Estimation Comparison

We performed an analysis of the cost comparison for uncertainty estimation for videos ranging from 1 to 2000. As depicted in Fig. 4, the initial cost of deep ensembles is substantially greater than that of other methods. The State-of-the-Art FacerDir method exhibits a slightly greater initial cost relative to the conventional MCD, MCDB, and TTA approaches. However, for a higher number of video estimations, the method proved to be substantially more efficient.

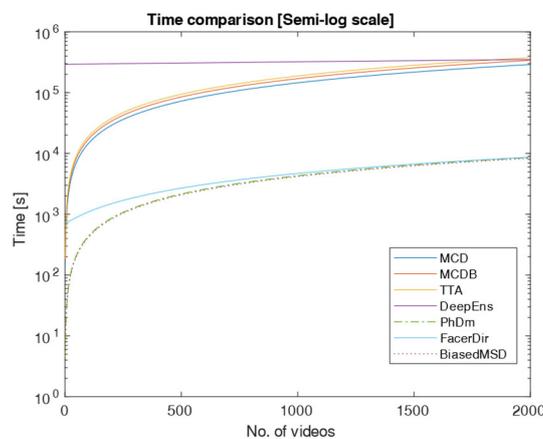


Fig. 4 Time comparison for uncertainty estimation for different methods. The plot represents the time required to estimate uncertainty using different methods over 2000 cumulative videos.

The novel proposed PhDm and BiasedMSD-based uncertainty estimation methods exhibit the least overhead time overall. The FacerDir, PhDm, and BiasedMSD-based uncertainty estimation methods exhibit considerably low overhead costs, indicating a promising opportunity to advance toward the subsequent phase of incorporating an active learning for data curation.

4.3 Baselines for Data Curation

Active learning is performed iteratively till the quarter of the informative/uncertain dataset is sampled. The model is later trained on the sampled informative training data and the following baselines are used for comparing the model performance:

4.3.1 Baseline for pool data curation

Pool data curation is one of the main purposes of active learning, while the most informative data is curated for annotation using a sampling strategy. The function $a(x)$ is defined as a draw from a uniform distribution over the interval $[0, 1]$ using the function (Gal, *et al.* 2017). This criterion ensures that the selection strategy for acquiring annotated data is superior to a random selection approach and is used in majority of scientific studies.

4.3.2 Baseline for train data curation

Active learning is seldom used in the literature to curate the training data as it is always seen as a method to curate the unknown pool data. In our research, an active learning approach was utilized to curate the training dataset with the objective of mitigating potential implicit biases that may exist within the data. The assessment and comparison of various approaches are based on the performance of the primary network, which is trained on the complete dataset utilized for active learning experimentation. In the event that the active learning technique produces a subset of the training dataset that surpasses the primary network's performance, it can be utilized for training data curation in a general context.

4.4 Active Learning for Data Curation

We conduct a series of experiments on three different data filter sizes viz. 15, 30, and 60 randomly filtered frames per sequence (fpseq). It is in-feasible to perform the experiments on the full internal dataset and hence, for research experimentation three different data filter sizes were randomly developed. It is important to note that the comparison of the metric over different data sizes internally is not possible as the data was randomly selected.

The uncertainty estimation was performed using PhDm, BiasedMSD, and FacerDir. The uncertainty estimation was iteratively used in an AL setting to sample top uncertain samples. The uncertain samples were iteratively added and removed from the training and pool dataset, respectively. The uncertainty sampling for PhDm and FacerDir-based approaches were done using Sequence-wise (SW) and Frame-wise (FW) querying strategies. Fig. 5 represents the comparison of the performance of different experiments over various data sizes.

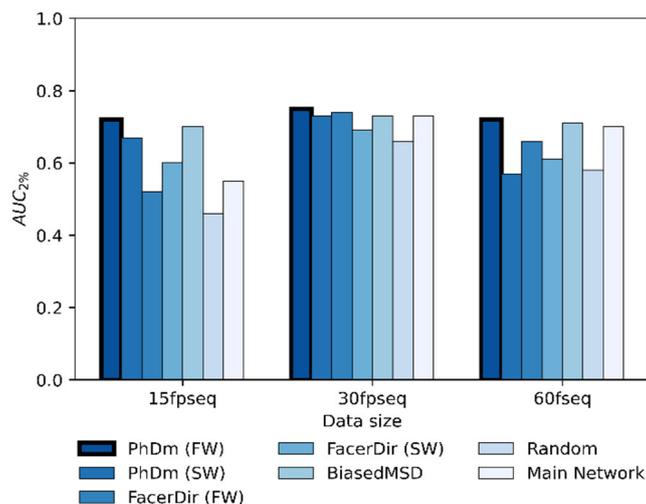


Fig. 5: AL method performance comparison for different methods applied over various filtered data sizes (bold bar represents best performing method). FW and SW represent Frame-wise and Sequence-wise querying strategies.

5. DISCUSSION

Fig. 4 illustrates the overhead cost associated with uncertainty quantification. It can be observed that the utilization of PCA and biased differencing-based techniques resulted in relatively shorter estimation times. This observation suggests that the proposed novel methods in the paper exhibit significantly lower overhead uncertainty estimation costs in comparison to the conventional and State-of-the-Art distillation approaches.

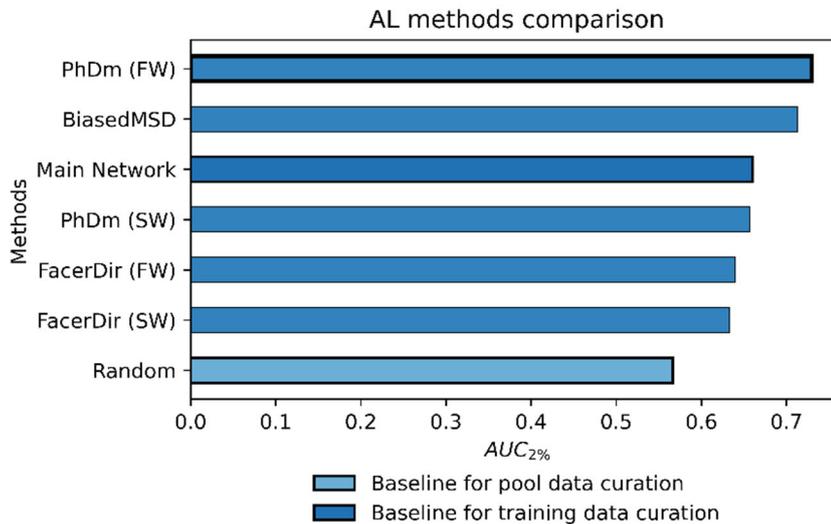


Fig. 6 Performance comparison of different AL methods. The random network is the baseline for pool data curation and the main network is the baseline for training data curation. The novel PhDm approach performance outperformed the random and main network baseline.

It is evident from Fig. 6 that the novel PCA-based method outperformed other AL approaches in the context of training and pool data curation. The overall comparison of different AL approaches is illustrated in Fig. 6. It is evident from the figure that the frame-wise selection approach was generally better than the sequence-wise querying strategy. The frame-wise selection of images from the videos proved to perform efficiently compared to selecting the whole sequence at a time using a sequence-wise selection strategy. The frame-wise querying strategy improved the performance of the PCA-based active learning method by 7.33% and the FacerDir approach by 0.67%. This finding can be utilized to crop informative frames from different excessively long videos, which is an intriguing application of the active learning strategy. In general, every method outperformed the Random baseline. This indicated that every method performed well in the context of pool data curation and annotating the pool data using uncertainty quantification.

In every experiment conducted using varying data sizes, the novel PhDm approach outperformed the main network, despite utilizing only 26.39% of the available data for model training. The approach improved the performance of the model architecture compared to the random baseline with 16.33% of AUC_{2%} evaluation metric. The approach even improved the main network performance by 7.33% and outperformed other approaches. These findings suggest that the PhDm approach can achieve superior performance with significantly fewer training data as compared to the main network and is an approach with a very low overhead uncertainty quantification cost.

BiasedMSD selects challenging data instances that the model struggles to comprehend as uncertain samples. Therefore, this approach is expected to yield superior uncertainty estimation results and excel in AL scenarios. By including image samples that the model has identified as false positives, the approach can effectively enhance data quality. The BiasedMSD approach improved main network performance by 5.66%, however, the improvement was not as evident as seen in the PhDm approach. The performance of FacerDir-based active learning was found to improve the network by 7.33% than the random baseline compared to a significant 16.33% improvement using the novel PCA-based hard data mining approach.

Recapitulating the results observed in the experiments, the feature-based novel approach outperformed other approaches in the context of curating pool and training data using AL. This is attributable to the fact that the feature-based approach captured more information about the samples than sampling a probability distribution over several forward passes.

6. CONCLUSION

This paper focuses on the various uncertainty estimation techniques for the object detection-based fire detection problem with an aim to curate the data. We proposed two novel approaches for curating the data using active learning. The novel PhDm and BiasedMSD approaches performed uncertainty estimation efficiently compared to the sampling-based methods and the State-of-the-Art FacerDir. The novel approaches were also found to outperform other benchmark methods in the task of curating the unlabeled data. PhDm was found to be the most efficient method for curating and improving the training data by decreasing the size of the training.

Finally, we put forward potential avenues for future research exploration. Different uncertainty and active learning experiments were performed in the setting of binary classification. The experimentation on multi-class classification problems can be done using the novel approaches stated in the literature. The novel AL techniques can be further evaluated on various publicly available datasets.

REFERENCES

- Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93-104). <https://doi.org/10.1145/342009.335388>
- Campbell, R. (2018). Fire in Industrial or Manufacturing Properties. Technical report, National Fire Protection Association (NFPA). Retrieved July 24, 2023, from NFPA report website: <https://www.nfpa.org/News-and-Research/Data-research-and-tools/Building-and-Life-Safety/Fires-in-US-Industrial-and-Manufacturing-Facilities>
- Choi, J., Elezi, I., Lee, H.-J., Farabet, C., & Alvarez, J. M. (2021). Active Learning for Deep Object Detection via Probabilistic Modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10264-10273). <https://doi.org/10.48550/arXiv.2103.16130>
- Deepshikha, K., Yelleni, S. H., Sriji, P. K., & Mohan, C. K. (2021). Monte Carlo DropBlock for Modelling Uncertainty in Object Detection. In *Computing Research Repository (CoRR)*. <https://doi.org/10.48550/arXiv.2108.03614>.
- Desai, S. V., Chandra, A. L., Guo, W., Ninomiya, S., & Balasubramanian, V. N. (2019, October). An Adaptive Supervision Framework for Active Learning in Object Detection. In *Proceedings of the British Machine Vision Conference*. <https://api.semanticscholar.org/CorpusID:199472921>
- Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International conference on machine learning, PMLR* (pp. 1050-1059). <https://proceedings.mlr.press/v48/gal16.html>
- Gal, Y., Islam, R., & Ghahramani, Z. (2017). Deep Bayesian Active Learning with Image Data. In *International conference on machine learning, PMLR* (pp. 1183-1192). <https://doi.org/10.48550/arXiv.1703.02910>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Advances in neural information processing systems*, 30. <https://dl.acm.org/doi/10.5555/3295222.3295309>
- Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Advances in neural information processing systems*, 30. <https://api.semanticscholar.org/CorpusID:6294674>
- Malinin, A., & Gales, M. (2018). Predictive Uncertainty Estimation via Prior Networks. *Advances in neural information processing systems*, 31. <https://doi.org/10.48550/arXiv.1802.10501>
- Malinin, A., Mlodozieniec, B., & Gales, M. (2019). Ensemble Distribution Distillation. In *8th International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1905.00076>

Manivannan, I. (2020). A Comparative Study of Uncertainty Estimation Methods in Deep Learning Based Classification Models. Technical report, Hochschule Bonn-Rhein-Sieg University of Applied Sciences, Department of Computer Science. <https://doi.org/10.18418/978-3-96043-085-8>

Nguyen, V.-L., Shaker, M. H., & Hüllermeier, E. (2022). How to measure uncertainty in uncertainty sampling for active learning. *Machine Learning*, 111(1), 89-122, <https://doi.org/10.1007/s10994-021-06003-9>

Schorn, C., & Gauerhof, L. (2020). FACER: A Universal Framework for Detecting Anomalous Operation of Deep Neural Networks. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems* (pp. 1-6). <https://doi.org/10.1109/ITSC45102.2020.9294226>

Sousa, M. J., Moutinho, A., & Almeida, M. (2020). Thermal Infrared Sensing for Near Real-Time Data-Driven Fire Detection and Monitoring Systems. *Sensors* (Basel, Switzerland), 20(23), 6803. <https://doi.org/10.3390/s20236803>

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, 52(2), 121-128. <https://doi.org/10.1177/070674370705200210>

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. (2015). Going Deeper with Convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9). <https://doi.org/10.1109/CVPR.2015.7298594>