Міністерство освіти і науки України

Національний університет «Києво-Могилянська академія»

Факультет інформатики

Кафедра математики

Магістерська робота

освітній ступінь – магістр

на тему: **«АКТИВНЕ НАВЧАННЯ У ЗАДАЧІ ВИЯВЛЕННЯ ОБ'ЄКТІВ»**

Виконала: студентка 2-го року навчання

освітньо-наукової програми «Прикладна математика»,

спеціальності 113 Прикладна математика

Ронська Дарина Романівна

Керівник: Швай Н. О.,

кандидат фіз.-мат. наук, доцент

Рецензент_____

(прізвище та ініціали)

Кваліфікаційна робота захищена з оцінкою _____

Секретар ЕК _____

«____» _____ 20____ р.

Київ – 2022

Table of Contents

Abstract

Active Learning allows to spend less time on data labeling which is vastly beneficial in Computer Vision with continuously growing number of datasets and its images. It is achieved by smarter strategy than random one to queue the images for labeling that allows to give most informative images to the model first. In this work state-of-the-art Multiple Instance Active Learning for Object Detection (MI-AOD) method is improved by the changes in its uncertainty function which corresponds for informativeness of the image. Also, the statement of MI-AOD authors about its usage for noisy images filtering is proved.

Introduction

It is said that at least three quarters of time in data science are spent on collecting and preprocessing data. That means that a significant amount of time is dedicated to datasets organization.

Labeling is one of the most time-consuming parts of the data organization, and despite of labeling is a routine process it involves a lot of human interaction. It can be hardly automized, because a supervised learning model needs a labeled data as an input and before the model is trained it has no predictive power at all. Another concern is an accuracy – the input for the model or the ground truth should be as much accurate as it is possible, not some kind of approximation.

Nevertheless, even though labeling process cannot be fully automized it can be made faster and easier for a human using active learning. Active learning is used to choose the most informative samples from the data lake [2]. Queueing unlabeled samples by their informativeness values helps to achieve better results in smaller number of samples labeled. So, the goal of active learning is to find the strategy to select samples for labeling better than random selection.

In this work some modifications are applied to the Multiple Instance Active Learning for Object Detection (MI-AOD) [1], which is state-of-the-art in active learning for object detection, to enhance the results. Also, hypothesis of filtering the blurred images using the model is checked, as this approach is claimed by the others to be able to suppress noisy instances.

In the first chapter the idea of active learning, object detection task and MI-AOD approach are explained.

The second chapter contains the experiments with the alternatives to instance uncertainty re-weighting metrics.

In the third chapter the application of the MI-AOD for the blurred images filtering is introduced and tested.

Chapter 1. Terms explanations and methods overview.

In this chapter main terms that are needed for the understanding of the work are explained. Also, state-of-the-art active learning methods for object detection are described here.

1.1. What is Object Detection?

Object Detection is related to the computer vision field. The task is to find the bounding boxes and their corresponding classes on the image. As an input model receives the image with one or more objects. As an output it gives one or more bounding boxes and class label for each bounding box. Usually, bounding boxes are defined by two points or as a point, width and height. These days, models from R-CNN and YOLO families are frequently used for Object Detection.

1.2. What is Active Learning?

Usually, pictures for the object detection are labeled by the people. With a large amount of data it might take a lot of time, costs and affords to label those pictures. That's why it can be helpful not to use all the pictures for the model, but to select the most informative ones for the model.

Active Learning is the subset of machine learning to choose samples for labeling in a smarter way. The main idea is to use the current knowledge of a model to select the most valuable samples for labeling, which would be a more effective way to improve the model results compared to a randomly chosen samples.

Various methods are used to choose the samples for the model, but usually algorithm for the active learning contains same main steps.

Let's assume, that there is unlabeled data pool U from which the data for labeling is takes, and labeled data L, where there are already some labeled samples (samples can be randomly selected and labeled from U for initial step).

1) Train the model with data from L.
2) Make predictions for the samples from U.

3) According to predictions in step 2 choose the most informative samples.
4) Label most informative samples from *U* and add them to *L*.
5) Repeat steps 1 - 4 until satisfactory results.

As it can be seen in the step 3, active learning requires some sort of informativeness measure for the unlabeled instances. It can be least confidence in its most likely label, or margin sampling - the difference between the first and the second most probable samples.

1.3. MI-AOD: Multiple Instance Active Learning for Object Detection

The main goal of the active learning for object detection is to choose the most informative images for the training of the detector model. However, most methods that are proposed for now tackle it by a simple summarizing or averaging the instances as image uncertainty. The main drawback of this naïve approaches is that all the instances on the images are treated equally, and it causes a large imbalance while calculating the average uncertainty, which is caused by the noisy instances in the background.

MI-AOD [1] approach targets at informative images selection from the unlabeled set by instance uncertainty re-weighting and multiple instance learning.

The MI-AOD initially defines an instance uncertainty learning (IUL) module that uses two adversarial instance classifiers inserted on top of the detection network (for example, a feature pyramid network) to learn the uncertainty of unlabeled instances in order to learn the instance-level uncertainty. While decreasing classifier discrepancy drives learning features to lessen the distribution bias between the labeled and unlabeled examples, maximization of the prediction discrepancy of two instance classifiers predicts instance uncertainty.

In addition to the instance classifiers, MI-AOD also includes a MIL module to establish the connection between instance and image uncertainty. MIL accomplishes instance uncertainty re-weighting (IUR) by assessing instance appearance consistency across images, treating each unlabeled image as an instance bag. A classification loss based on picture class labels is forced to consistently drive the instance uncertainty and image uncertainty during MIL (or pseudo-labels). Suppressing the noisy instances while

emphasizing the truly representative ones is made easier by optimizing the image-level classification loss. In order to choose the most informative images for detector training, instance-level observation and image-level evaluation are linked by iterative instance uncertainty learning and instance uncertainty re-weighting.

Chapter 2. MI-AOD instance uncertainty metric modifications

In this chapter, modifications of uncertainty re-weighting metric are proposed and tested.

2.1. MI-AOD instance uncertainty original metric

RetinaNet is used as the baseline. A detector with two discrepant instance classifiers (f_1 and f_2) and bounding box regressor f_r , parametrized by θ_{f_1} , θ_{f_2} and θ_{f_r} correspondingly. The prediction discrepancy between the two instance classifiers is utilized to learn the instance uncertainty on the unlabeled set. The feature extractor g is parametrized by θ_g. $\Theta = \{\theta_{f_1}, \theta_{f_2}, \theta_{f_r}, \theta_g\}$ denotes the set of all parameters. θ_{f_1} and θ_{f_2} are initialized independently. x can be represented by multiple instances $\{x_i,\ i = 1, \dots, N\}$ corresponding to feature anchors on the feature map. N is the number of the instances in image x. $\{y_i,\ i = 1, \dots, N\}$ denote the labels for the instances.

There is a distribution bias between the labeled and unlabeled set before the labeled set can accurately reflect the unlabeled set, especially when the labeled set is small. The biased distribution area contains the informative examples. f_1 and f_2 are developed as adversarial instance classifiers with greater prediction discrepancy on the instances near the boundary in order to identify them. The difference between f_1 and f_2 predictions is what is referred to as the instance uncertainty.

In this process, θ_g is fixed which guaranties distributions of both labeled and unlabeled instances are fixed. At the same time, prediction discrepancies for all instances are maximized on the unlabeled set while the detection performance on the labeled set is preserved. For these the following loss function is optimized:

$$\underset{\Theta \setminus \theta_g}{\operatorname{argmin}} \mathcal{L}_{max} = \sum_{x \in \mathrm{X}_L} l_{det}(x) - \sum_{x \in \mathrm{X}_U} \lambda \cdot l_{dis}(x),$$

where

$$l_{dis}(x) = SSE(x) = \sum_i \left(\hat{y}_i^{f_1} - \hat{y}_i^{f_2}\right)^2$$

$l_{dis}(x)$ denotes the prediction discrepancy loss and $l_{det}(x)$ is a detection loss. $\hat{y}_i^{f_1}$, $\hat{y}_i^{f_2} \in \mathbb{R}^{1\times C}$ are the instance classification predictions of the two classifiers for the i-th instance in image x, where C is the number of object classes in the dataset, and λ is a regularization hyper-parameter determined by experiment.

2.2. Proposed metrics for uncertainty re-weighting

In this chapter prediction discrepancy loss l_{dis} is modified and the performance of the model is evaluated by mean average precision (*mAP*).

2.2.1. Binary Cross Entropy

Cross-entropy [3] is a measure borrowed from the field of information theory, based on entropy. It calculates the difference between two probability distributions:

$$BCE(x) = BCE\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = -\sum_i \hat{y}_i^{f_1} \cdot log(\hat{y}_i^{f_2}) + (1 - \hat{y}_i^{f_1}) \cdot log(1 - \hat{y}_i^{f_2})$$

2.2.2. Focal Loss

Focal loss [4] applies a modulating terms to the cross-entropy loss in order to focus learning on hard misclassified examples. α and γ are hyperparameters that are used for further calibration of the model. More the value of γ, more importance will be given to misclassified examples and very less loss will be propagated from easy examples. α term is used to handle both the foreground and background class imbalance and hard negative samples' gradient salience. α and γ are usually set to 0.25 and 2 correspondingly.

$$FL(x) = FL\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = \sum_i \mathrm{A}_i(x) \cdot \Gamma_i(x) \cdot CE_i(x),$$

where

$$\mathrm{A}(x) = \mathrm{A}\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = \alpha\hat{y}^{f_1} + (1-\alpha)\left(1 - \hat{y}^{f_1}\right),$$

$$\Gamma(x) = \Gamma\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = \left(1 - \left(\hat{y}^{f_1}\hat{y}^{f_2} + \left(1 - \hat{y}^{f_1}\right)\left(1 - \hat{y}^{f_2}\right)\right)\right)^{\gamma},$$

$$CE(x) = CE\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = -\hat{y}^{f_1}\, log(\hat{y}^{f_2}) + (1 - \hat{y}^{f_1}) \cdot\, log(1 - \hat{y}^{f_2}),$$

$$\alpha = const, \qquad \gamma = const.$$

2.2.3. Kullback–Leibler (KL) divergence

It is closely related to but is different from cross-entropy that calculates the total entropy between two probability distributions, whereas KL divergence [5][6] can be thought to calculate the relative entropy between the distributions.

$$KL(x) = KL\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = \sum_i \hat{y}_i^{f_1} \cdot log\left(\frac{\hat{y}_i^{f_1}}{\hat{y}_i^{f_2}}\right)$$

2.2.4. Jensen-Shannon (JS) divergence

Jensen-Snannon Divergence [7] is based on the Kullback–Leibler divergence, with some notable (and useful) differences, including that it is symmetric and it always has a finite value.

$$JS(x) = JS\left(\hat{y}^{f_1}, \hat{y}^{f_2}\right) = \frac{1}{2} KL\left(\hat{y}^{f_1}, \frac{\hat{y}^{f_1} + \hat{y}^{f_2}}{2}\right) + \frac{1}{2} KL\left(\hat{y}^{f_2}, \frac{\hat{y}^{f_1} + \hat{y}^{f_2}}{2}\right)$$

2.3. Metrics comparison and results

Both Binary Cross Entropy and Jensen-Shannon Divergence show slightly better performance than the original metric *(Figure 2.1)*. These metrics use probability distributions to find difference between them, while L2 norm is not precisely for probability distributions, but for any continuous ones. Binary Cross Entropy gives better results than L2 norm starting from the first non-random images selection. On the other hand, Jensen-Shannon Divergence shows the best result on the last images selection and good overall results on intermediate images selections. Kullback-Leibler Divergence shows worse results than Jensen-Shannon Divergence probably because

this metric is not symmetrical, which is the main difference between these two divergences.

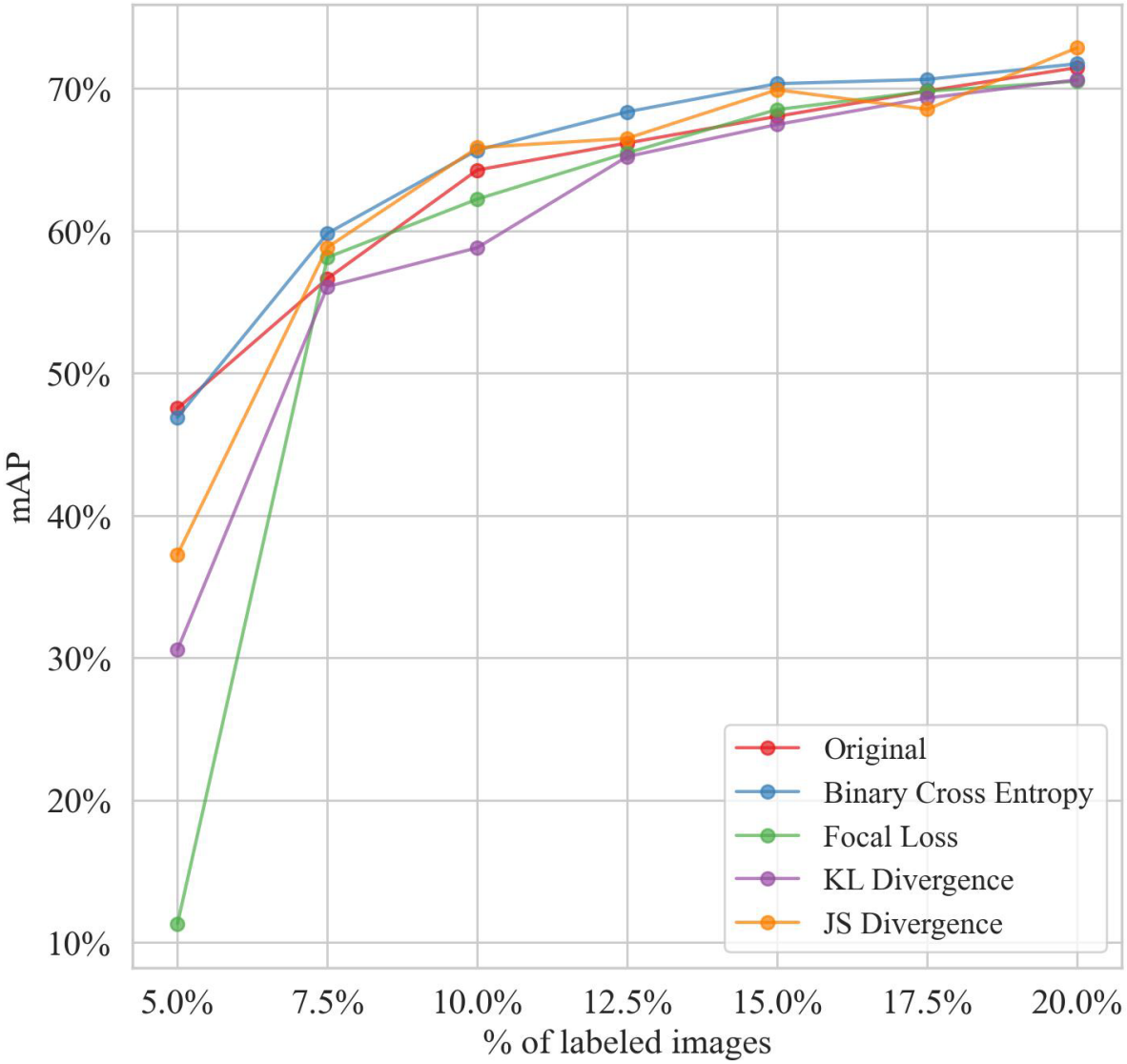

Figure 2.1. Performance comparison for different uncertainty metrics

Chapter 3. MI-AOD application for blurred pictures filtering

3.1. Motivation to use MI-AOD for blurred pictures filtering

In the paper [1] it is said that "*instance uncertainty learning (IUL) and instance uncertainty re-weighting (IUR) modules, providing effective approaches to highlight informative instances while filtering out noisy ones in object detection.*" This statement on its own was not tested by the authors. That is why experiment to testify the statement is provided in this work.

So, for this experiment images from both train and test set were used. For half of the images Gaussian blur was applied and the other half of the images was remaining unchanged.

3.2. Gaussian blur

The Gaussian blur [8] functionality is obtained by blurring (smoothing) an image using a Gaussian function to minimize the noise level. It can be thought of as a nonuniform low-pass filter that maintains low spatial frequency while reducing image noise and minor information. It's usually done by using a Gaussian kernel to convolve a picture.

This Gaussian kernel in 2-D form is expressed as:

$$G_{2D}(x, y, \theta) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

where σ is the standard deviation of distribution and x and y are the position indices. The value σ of determines the magnitude of the blurring effect around a pixel by controlling the variation around a mean value of the Gaussian distribution.

3.3. Blurred image classification by uncertainty

After half of the images from the test set were blurred, for each image uncertainty was calculated using the models in the original version of MI-AOD.

The distributions of uncertainty for blurred and original images are looking different for both train and test sets *(Figure 3.1)*.

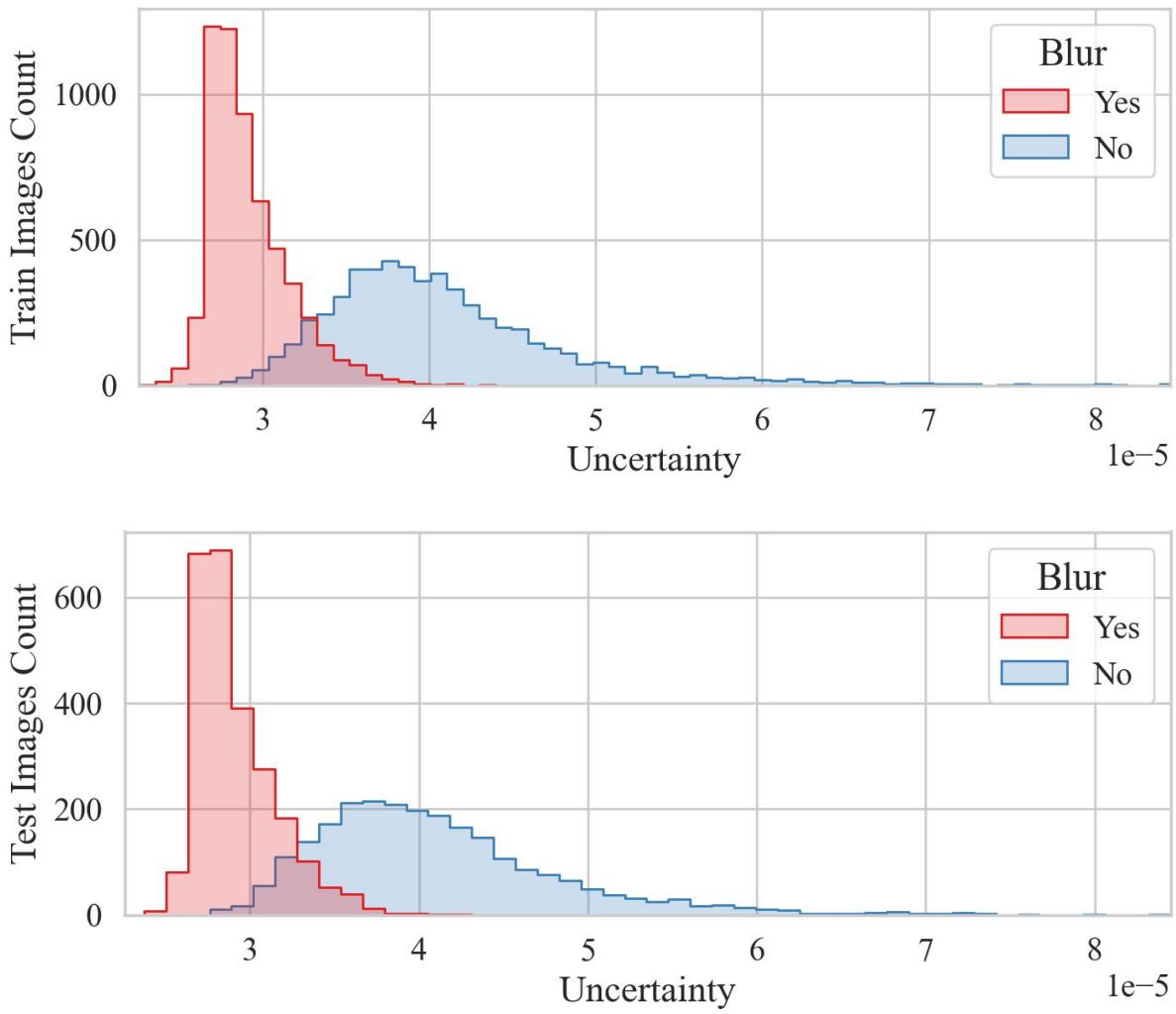

Figure 3.1. Histograms of uncertainty distributions among blurred and clear images for train and test sets

Precision-recall curve of the train set is used to find the threshold to define the class for each image. Threshold is defined as the closest point to (1, 1) on the precision-recall curve *(Figure 3.2)*.

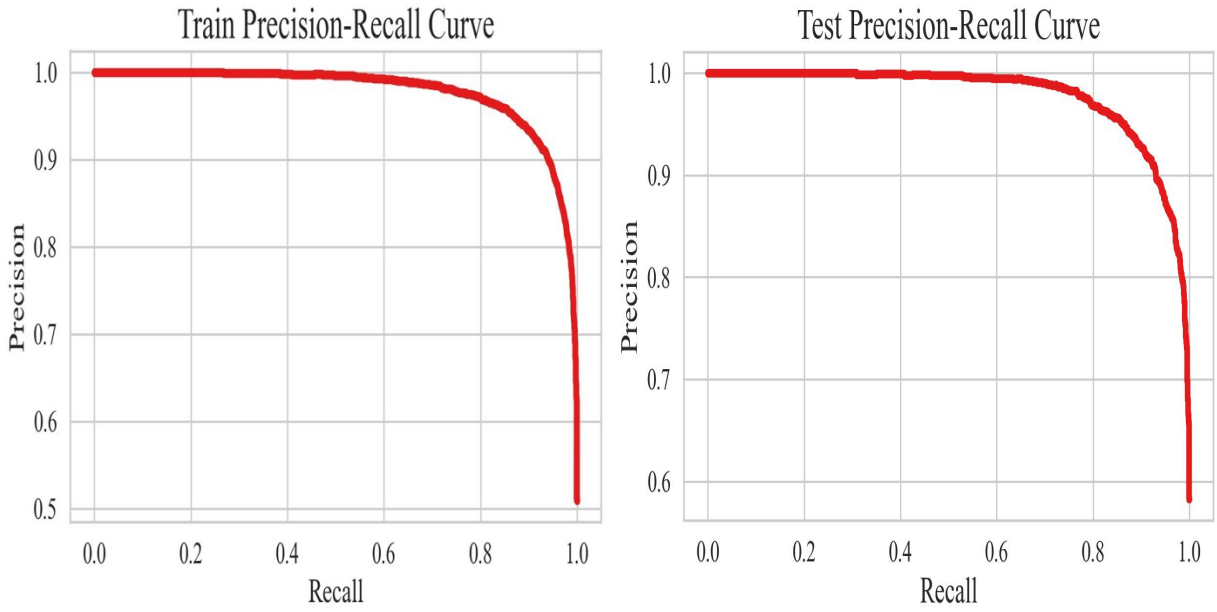

Figure 3.2. Precision-recall curves for blur classified images for train and test sets

The accuracy on the train set is 92% and 91% on the test set *(Table 3.1)*.

Table 3.1. Blur classification metrics comparison for train and test sets

Metric	Train	Test
Accuracy	0.921	0.914
Precision	0.910	0.899
Recall	0.935	0.931
F1-score	0.922	0.914

Overall results are good among all metrics and both train and test sets, which shows stable performance and no signs of overfitting. Confusion matrix *(Figure 3.3)* shows, that the performance of the model is quite balanced among classes.

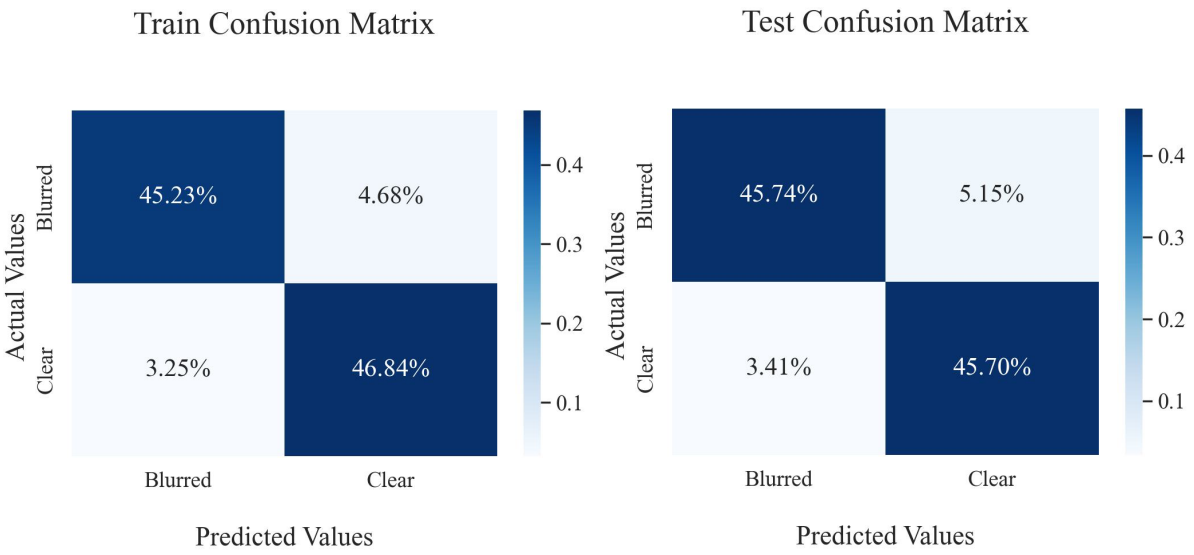

Figure 3.3. Confusion matrices for blur classified images for train and test sets

Summary

Dataset for object detection tasks takes a lot of time to be prepared. Labeling is an essential part of data preparation for any type of supervised learning tasks. Thus, it cannot be fully automized, active learning may help to spend less time on this, which also means less costs and work on the labeling.

In the first part of this work method MI-AOD, which shows one of the best results among other active learning methods for object detection, was improved. It is achieved by modification of the uncertainty function which is further used in the model loss function. Originally, in MI-AOD L2-norm was used as an uncertainty function. In our experiments - Entropy, Focal Loss, Jensen-Shannon divergence and Kullback-Leibler divergence metrics were tried to substitute the original function. These metrics were chosen as they were developed to find the divergence between the two probability distributions, which in theory satisfies the task to find the uncertainty more than L2-norm. The model trained with Cross-Entropy uncertainty function has stable and better performance than the original one – it gives a 1-3% boost on each iteration (images selection step) compared to L2-norm model.

The other part of this work is devoted to proof of the statement which was given in the MI-AOD paper by its authors. There they say that the models that are trained using MI-AOD approach can be used to filter noisy images. To prove this, some Gaussian blur was added on the half of the images in the dataset and uncertainty was calculated for all the images. The distribution of uncertainty for blurred and clear images differed a lot, which could be easily seen on the histograms. After threshold for classes split was calculated from train set uncertainty distributions, it was used to estimate the results of the blur images classification. The accuracy of this classification on the train and test set is 92% and 91% correspondingly. Other metrics such as precision, recall and f1-score were calculated and showing good results as well as no signs of overfitting.

So, in this work the method MI-AOD was improved by the uncertainty metric modification and the statement of the MI-AOD authors about usage of the models for noisy images filtering was justified.

References

1. Yuan, Tianning, et al. "Multiple instance active learning for object detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
2. Settles, Burr. "Active learning literature survey." (2009).
3. Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
4. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.
5. Kullback, Solomon. "Information theory and statistics. john riley and sons." *Inc. New York* (1959).
6. Kullback, S., and R. A. Leibler. "10.1214/aoms/1177729694." *Ann. Math. Stat* 22 (1951): 79-86.
7. Nielsen, Frank. "On the Jensen–Shannon symmetrization of distances relying on abstract means." *Entropy* 21.5 (2019): 485.
8. Shapiro, L. G., and G. Stockman. "C:" Computer Vision", page 137, 150. Prentice Hall." (2001).