Evidential Query-By-Committee Active Learning for Pedestrian Detection in High-Density Crowds

Jennifer Vandoni, Emanuel Aldea, Sylvie Le Hégarat-Mascle

SATIE laboratory, University Paris-Sud, 91405 Orsay cedex, France

Abstract

The automatic detection of pedestrians in dense crowds has become recently a very active topic of research due to the implications for public safety, and also due to the increased frequency of large scale social events. The detection task is complicated by multiple factors such as strong occlusions, high homogeneity, small target size, etc., and different types of detectors are able to provide complementary interpretations of the input data, with varying individual levels of performance. Our first contribution consists in outlining a fusion strategy under the form of an ensemble method, which models the imprecision arising from each of the detectors, both in the calibration and in the spatial domains in an evidential framework. Then, we propose a sample selection for augmenting the training set used jointly by the committee of classifiers, based on evidential disagreement measures among the base members in a Query-by-Committee context. The results show that the proposed fusion algorithm is effective in exploiting the strengths of the individual classifiers, as well as in augmenting the training set with informative samples which allow the resulting detector to enhance its performance.

Keywords: Pedestrian detection, Crowd analysis, Ensemble methods, Belief function theory, Active learning

1. Introduction

For video surveillance, the automatic detection of pedestrians is a fundamental task which is directly related to applications such as tracking or action recognition. The context of the detection application may range widely, with works addressing various topics such as safety issues for autonomous driving [1, 2], fall detection for elderly

people [3], surveillance [4] or automated video anonymization [5]. Recently, the accurate detection of pedestrians in high-density scenes gained traction due to the increased frequency of large scale social events, and due to the safety risks linked to them [6]. Although a significant effort has been devoted in the last decade to pedestrian detection [7, 8], the advances proposed in the literature are not always applicable to high-density crowd detections for multiple reasons [9], such as the difficulty to obtain an adequate training set and the intrinsic complexity of the scenes.

Firstly, common pedestrian detectors (e.g. [10, 11]) are trained for discriminative learning on the basis of a large labeled training set. In case of extremely dense crowds however, it may become hard to define a good training set which spans over all the possible shades of sample characteristics while at the same time remaining focused on the specific targets. To this extent, Active Learning (AL) has been proposed [12]. It relies on the assumption that if a learning algorithm is allowed to choose data from which to learn, it will reach better levels of performance with less training data [13]. Secondly, common problems for the detection task in high-density crowds are the absence of background, the heavy occlusion of body parts, the high visual homogeneity and the small size of the targets. It becomes therefore essential to rely on multiple independent visual detectors which are able to provide different interpretations of the input data. However, it is not immediately clear which detectors are the most adapted or discriminative, and which fusion strategy is the most effective to get the best out of their combination.

Active learning background:. Pool-based AL [14] relies on an initial small set of labeled instances, $\mathcal{L}$, and a larger set of unlabeled ones, $\mathcal{U}$. Batches of *informative* training samples are iteratively selected from $\mathcal{U}$ and added to $\mathcal{L}$, with respect to some heuristics, after a query about their actual label to an *oracle* (e.g., a human annotator). This approach is well-motivated in many modern machine learning applications, where unlabeled data may be abundant but labels are difficult, time-consuming, or expensive to obtain, from text classification [15] to robotics [16] and medical image classification [17] among others.

Many strategies to select new training samples have been proposed. The most pop-

ular ones are uncertainty sampling and Query by Committee (QBC), with many variations in order to balance exploitation of the current classifiers and exploration of the version spaces [18]. Uncertainty sampling [19] consists in iteratively requesting labels for training instances whose classes remain uncertain, despite the information provided by the previously labeled instances. In this way the learning algorithm can focus its attention on the examples it finds confusing, selectively adjusting the boundary between classes. Popular strategies consist in querying the instance whose predicted output is the least confident or with maximum entropy, but in the context of SVM classification the prevailing method is to select the samples which are closer to the separation hyperplane margin [20, 21]. More recently, DUAL [22] and QUIRE [23] methods have been proposed. The former is based on density weighted uncertainty sampling while the latter aims at selecting both informative and representative examples on the basis of a prediction of the uncertainty. The authors of [24] consider instead the *diversity* between samples, proposing a selection strategy which aims to reach a trade-off between the minimum distance from the hyperplane margin and the maximum angle between the hyperplanes defined by each sample. In the context of image classification, diversity among the selected samples can be reached using spatial information, such as in [25], where the authors propose three criteria to favor samples distant from the ones already present in the training set, namely an Euclidean distance, a distance based on the Parzen window method applied in the spatial domain and a distance that maximizes the spatial entropy variation value to distribute spatially the training samples as widely as possible. Although uncertainty sampling offers an intuitive and flexible solution for augmenting the training set, this framework is suited in its standard form for relying on a single classifier.

On the other hand, QBC [26] exploits a committee of classifiers and operates by asking for the label of the sample on which the ensemble disagrees the most. This approach is better suited for more complex classification tasks which benefit from multiple classifiers providing different views of the input data, such as the application we consider here. Three questions arise, namely how to build the committee set, how to quantify the disagreement in order to define a strategy to select the new samples, and how to finally combine the committee member responses in order to obtain a robust

classification. Usually generic ensemble learning algorithms are used for the construction of the committee. Query-by-bagging [27] or query-by-boosting [28] can be used to train weak classifiers on (weighted) randomly sampled variations of the training data set. Alternatively, a single model can be exploited and many variations of it can be derived, e.g. changing its intrinsic parameters, like in [29] for naive Bayes, using the Dirichlet distribution over model parameters.

There exists a variety of heuristics to measure the disagreement among a classifier ensemble, but surely the most popular ones are (Soft) Vote Entropy [13], and Kullback-Leibler (KL) divergence [30]. Other measures include Jensen-Shannon divergence [31], a smoothed version of KL divergence, and F-compliment [32], based on the F1-measure. A combination between Vote Entropy and KL divergence is proposed in [33] in the specific context of stream-based QBC, where a continuous stream of samples is given as input and the active learner must decide if it is worth or not asking for the true label. Recently, [34] proposed an interesting method to incorporate diversity and density measures in the instance selection, to ensure variety within the batch and in the whole training set.

Finally, the classification in the context of QBC is usually performed at every iteration on the basis of the committee member responses, through an average among them (weighted, in case of boosting), or by picking the model that provides the best performance (e.g., accuracy). In cases like our application, where the committee is quite heterogeneous and there is not clearly an absolutely best classifier, but rather each classifier is independent and contributes providing a different view of the same data, a simple average between them may not exploit all the available information, so that an adapted fusion becomes an integral part of the entire process.

Classifier combination. In order to perform the fusion of detectors based on different features, there exist in the literature various approaches, more or less suited for pedestrian detection. To benefit simultaneously from all the available features, multiple kernel learning (MKL) is a well established methodology which aims to combine different kernels relying on different data representations as a linear combination, by casting this information fusion task as a convex optimization problem [35]. The prob-

lem scales very well with the number of individual classifiers, but the main limitation of MKL is the difficulty to interpret the final decision and to take into account the imprecision coming from different sources. Another established framework able to benefit from the information provided by multiple features is the decision tree analysis. Recent work highlighted that intrinsic uncertainty related to learning as well as uncertainty due to imprecise data may be jointly managed inside the decision tree by defining entropy intervals from evidential likelihood [36]. In [37] instead, a set of base classifiers is dynamically selected for each test sample on the basis of a classification gain computed using a probabilistic model that uses the outcome from previous observations. Information gain is employed also in [38] for actively selecting features combining the collected evidence over time while taking into account the amount of available training data for each class.

In the context of high-density crowd pedestrian detection, in [39] we proposed a robust fusion strategy based on the Belief Functions (BF) framework [40, 41, 42]. The evidential framework [43, 44, 36, 45] is indeed able to naturally model the concept of *imprecision*, that in our case can arise in two different and complementary ways: in the derivation of posterior probability values from SVM decision scores, and later, from the spatial layout of the detections in the output image space. The proposed mass allocation has been shown to be robust to possible imprecision of the calibration functions, while at the same time taking into account the information coming from neighboring pixels in the image space. Besides, it allows for an amount of discounting that is different at every pixel of the classifier's output map, and it is not only a constant value that merely reflects the reliability of the detector.

In the context of AL, a clear limitation of traditional QBC approaches however is that the selection of the new samples to be added to the training set is performed independently from the (optional) committee member combination, that is only used to derive statistics for evaluation purposes. The possible information arising from the combination of the committee members is not exploited. From our part, the definition of a fusion strategy based on BF framework allows us to naturally have at our disposal several clues to quantify the disagreement between committee members. The result of the source combination indeed is a *basic belief assignment* (bba) associated to every

unlabeled sample, that intrinsically contains conflict and ignorance components. For this reason we find it appropriate to work in the evidential domain: from the one hand, through the definition of appropriate bbas we can model the imprecision over the uncertainty value provided by each classifier; on the other hand, the BF framework directly provides indicators to quantify the disagreement between committee members.

In this study, we thus propose a QBC algorithm that takes a committee of models which are all labeled on the same training set, but representing competing hypotheses supported by different SVM classifiers based on gradient, texture and orientation descriptors. Firstly we use BF framework to perform fusion between the different pedestrian detectors, and then we propose and investigate different evidential-based measures for the selection of the batch of new training samples. The evidential framework is therefore not only involved in the combination of the sources to obtain a robust decision, but it plays at the same time an original role in the definition of new sample selection strategies at each iteration.

The contribution of this paper is twofold. Firstly, we define an evidential framework to perform active learning in a QBC context, based on the output bba obtained after the fusion of different classifiers. Secondly, we investigate the different evidential entropy definitions of the literature to this extent. The BF framework plays a key role in both cases, i.e., during the bba allocation and combination, and as input to derive evidential measures to select the new samples at every iteration of the active learning procedure.

In the following sections, we firstly explain the bba allocation and combination (Section 2), before arriving to the definition of the new strategies to select the samples based on evidential functions (Section 3). The experimental part (Section 4) illustrates, for our high-density pedestrian detection application, the impact of the different sampling strategies we considered on the performance metrics of the resulting detectors. Then, we conclude our study in Section 5.

2. Modeling classifier imprecision

2.1. Belief function framework

To handle both uncertainty and imprecision, belief functions are defined on a larger hypothesis set than in the case of the probabilistic framework. Specifically, if Θ denotes the discernment frame, i.e. the set of mutually exclusive hypotheses, belief functions are defined on the set of the subsets of Θ, noted 2^{Θ} in reference to its number of elements: $2^{|\Theta|}$ where $|\Theta|$ is the cardinality of Θ.

In our case, denoting by H and $\overline{H}$ the two singleton hypotheses,*"Head"* and *"Not Head"*, the discernment frame is $\Theta = \left\{H, \overline{H}\right\}$, and the set of hypotheses is $2^{\Theta} = \left\{\emptyset, H, \overline{H}, \left\{H, \overline{H}\right\}\right\}$.

Classically, the *mass* function noted m is the *basic belief assignment* (bba) that satisfies $\forall A \in 2^{\Theta}, m(A) \in [0,1]$, $\sum_{A \in 2^{\Theta}} m(A) = 1$. The hypotheses for which the mass function is non null are called *focal elements*. Then, other BF are in one-to-one relationship with m. They are used either for decision, namely the *plausibility* and the *credibility* functions noted Pl and Bel respectively, or for some computations. In this particular setting in which we have only two singleton hypotheses and $m(\emptyset) = 0$, Pl and Bel are defined by: $\forall A \in \left\{H, \overline{H}\right\}$ $Bel(A) = m(A)$ and $Pl(A) = m(A) + m(\Theta)$.

It is important to notice that Pl and Bel functions may also be interpreted as upper and lower probabilities [40] and they check the duality property: $\forall A \in 2^{\Theta}, Pl(A) = 1 - Bel(\overline{A})$ (where $\overline{A}$ denotes the complement of A with respect to Θ).

2.2. Bba definition based on calibrated scores

In the context of SVM-based high density crowds pedestrian detection, we consider that *imprecision* can arise in two different and complementary ways: in the derivation of posterior probability values from SVM decision scores, and later, from the spatial layout of the detections in the output image space. More specifically, let us explain better the origin of these two different types of imprecision.

Firstly, in order to obtain class probabilities from SVM scores, i.e. sample distances to the hyperplane margin, a well established method proposed by Platt [46] consists in

approximating the posterior probability by learning the optimal parameters configuration of a logistic sigmoid function, relying on a calibration set independent from the training data.

In particular, given training samples $x_j \in \mathbb{R}^n$, $j = 1, ..., l$, labeled by $y_j \in \{+1, -1\}$, defined as feature vectors derived from a head detector, the binary SVM computes a decision function $f(x)$ such that $sign(f(x))$ is used to predict the label of unseen test samples. In order to obtain class probability $P(y = 1|x)$, the method proposed by Platt [46] approximates the posterior probability by learning a logistic sigmoid function

$$P(y = 1|x) \approx \sigma_{\lambda_0,\lambda_1}(f) = \frac{1}{1 + e^{\lambda_0 f + \lambda_1}}. \quad (1)$$

The optimal parameter configuration $(\lambda_0^*, \lambda_1^*)$ is then determined by solving a regularized maximum likelihood problem, with respect to the calibration set.

For each different test sample, given its score s_i, namely its distance to the hyperplane boundary defined by classifier i, with $i = 1 \dots N$, we now define an associated Bayesian bba $m_i^{\mathcal{B}}$ (i.e., bba having only singleton focal elements), from the posterior probability given by the calibration step:

$$\begin{aligned} m_i^{\mathcal{B}}(H) &= \sigma_{\lambda_0^*,\lambda_1^*}(s_i), \\ m_i^{\mathcal{B}}(\overline{H}) &= 1 - \sigma_{\lambda_0^*,\lambda_1^*}(s_i), \\ m_i^{\mathcal{B}}(\Theta) &= 0, \\ m_i^{\mathcal{B}}(\emptyset) &= 0. \end{aligned} \quad (2)$$

This initial Bayesian bba is only able to model the uncertainty about the class the sample belongs to, relying on a calibration procedure that is assumed to be robust.

However, in difficult settings such as our application, a robust estimation of the sigmoid parameters is almost impossible to achieve, and few changes in the calibration set (cardinality or in the samples within it) can cause the sigmoid to appear very different. In presence of a steep transition between the two classes particularly, even a slight shift of the sigmoid may induce very different probability values and possibly different deci-

sions for quite numerous samples, especially in presence of strong overlap between the two classes. Now, with belief functions we can naturally take into account the imprecision inherent to the sigmoid learning process. Instead of deriving a simple probabilistic value through logistic regression, we aim at associating a bba to each unlabeled sample directly from its score and from the estimated sigmoid (from calibration process).

Xu et al. [47] proposed to extend the logistic calibration to derive a bba that takes into account the number of samples per score value for calibration process. Such an approach is suitable specially when the number of samples if small and when there is no overlapping between the scores of the two considered classes. Otherwise, [45] shows that, in such difficult types of applications, it is hard for SVM to find a very large margin between the two classes and there can be a consistent overlap between samples with different labels for the same score. However, since the number of samples per score would be high, we would paradoxically not assign a high value of imprecision to them.

Then, as an alternative we consider the bba allocation proposed by [48]. It relies on the observation that (fuzzy) erosion and dilation (respectively opening and closing) are also dual with respect to complementation, and they can be interpreted as belief and plausibility functions: given a bba m_0 derived from the output of a classifier, the following property holds:

$$Pl(A) = 1 - Bel(\overline{A}) \quad \leftrightarrow \quad \delta_v(m_0(A)) = 1 - \mathcal{E}_v(m_0(\overline{A})), \tag{3}$$

$$\leftrightarrow \quad \phi_v(m_0(A)) = 1 - \gamma_v(m_0(\overline{A})), \tag{4}$$

$\forall A \in 2^{\Theta}$, where δ_v and $\mathcal{E}_v$ are the dilation and erosion operators respectively, with structuring element v, while ϕ_v and γ_v are the closing and opening operators.

The amount and shape of the possible imprecision is thus modeled through a structuring element. Now, we propose to interpret the erosion operator as a discounting operator, in the sense that the obtained bba will be less committed. Indeed, when applying erosion to $m_0(A)$ to derive $Bel(A), \forall A \in \left\{H, \overline{H}\right\}$, the mass on Θ is increased by the sum of the differences between initial values and eroded values: $m_0(A) - \mathcal{E}_v(m_0(A)) + m_0(\overline{A}) - \mathcal{E}_v(m_0(\overline{A}))$.

In our case, the initial (Bayesian) bbas $m_i^{\mathcal{B}}$ are provided by the learned sigmoid

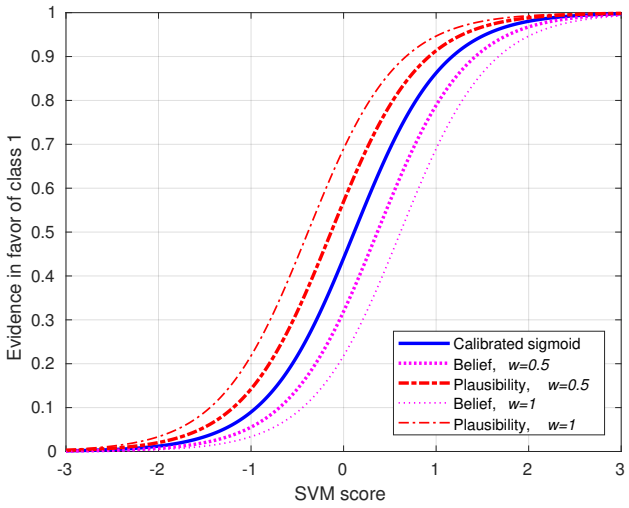


Figure 1: Example of a sigmoid function obtained with calibration, and derived Belief and Plausibility bounds at different structuring element w sizes. In our case, class 1 corresponds to the H hypothesis.

associated to each classifier *i*, through the probabilistic calibration.

Then, applying erosion and dilation operations to this sigmoid, with a structuring element of width w defined as a segment line in the score domain, allows for the derivation of two new sigmoid functions that are interpreted as lower and upper bounds of probability with respect to the learned sigmoid, i.e. Bel and Pl functions of the obtained bba. Due to the fact that we consider a flat structuring element and to the intrinsic monotonically increasing profile of the sigmoid function, considering classifier i, it is possible to easily derive:

$$Bel_i(H) = \sigma_{\lambda_0^*,\lambda_1^*}(s_i - \frac{w}{2}), \quad (5)$$

$$Pl_i(H) = \sigma_{\lambda_0^*,\lambda_1^*}(s_i + \frac{w}{2}). \quad (6)$$

Figure 1 shows an example of a sigmoid function learned on the calibration set, as well as the two derived sigmoid functions (for two structuring elements of different widths), that represent Bel and Pl functions and provide the interval of imprecision.

The interval between Bel and Pl functions embeds thus the amount of imprecision in the calibration step we have to cope with. It takes low values for points far from

Table 1: Example of bba allocation based on calibrated scores, assuming $\lambda_0^* = -2$, $\lambda_1^* = -0.05$ and erosion structuring element of width $w = 1$. Only the focal elements are reported.

Score	**Bayesian bba**	**Imprecise score-based bba**
$s_{x_1} = -0.5$	$m^{\mathcal{B}}_{x_1}(H) = 0.28$ $m^{\mathcal{B}}_{x_1}(\overline{H}) = 0.72$	$\widetilde{m}_{x_1}(H) = 0.12$ $\widetilde{m}_{x_1}(\overline{H}) = 0.49$ $\widetilde{m}_{x_1}(\Theta) = 0.39$
$s_{x_2} = +2$	$m^{\mathcal{B}}_{x_2}(H) = 0.98$ $m^{\mathcal{B}}_{x_2}(\overline{H}) = 0.02$	$\widetilde{m}_{x_2}(H) = 0.95$ $\widetilde{m}_{x_2}(\overline{H}) = 0.01$ $\widetilde{m}_{x_2}(\Theta) = 0.04$

the hyperplane boundary for which the decision is already pretty sure, whereas on the contrary it takes high values in the area near to the hyperplane margin, where even a slight difference in the parameters of the sigmoid can change the decision. Then, previous bba allocation allows us to model the fact that the calibration function may be not perfectly fitted due to the difficulty in the definition of a robust calibration set and to allocate large values of imprecision to the samples having their correspondent score within the SVM margin, in the overlapping area.

Table 1 proposes a toy example to illustrate the considered bba allocation based on SVM scores. Let us suppose that for a given classifier the sigmoid's optimal parameters have been found to be $\lambda_0^* = -2$ and $\lambda_1^* = -0.05$ through Platt's calibration based on logistic regression on the calibration set. Then, considering two different test samples x_1 and x_2, such that $s_{x_1} = -0.5$ and $s_{x_2} = +2$ are their SVM scores (i.e. their distances to the classification hyperplane), Eq. (1) provides the probability estimates $P(y = 1|x_1) = 0.28$ and $P(y = 1|x_2) = 0.98$. Then, we can derive the associated Bayesian bbas by simply assigning the probability estimate to the mass on H, and by computing the mass on $\overline{H}$ accordingly. For example, considering sample x_1, $m^{\mathcal{B}}_{x_1}(H) = P(y = 1|x_1)$ and $m^{\mathcal{B}}_{x_1}(\overline{H}) = 1 - P(y = 1|x_1)$. Then, by applying erosion with a flat structuring element of width w (in the example $w = 1$) we can discount the mass on singleton hypotheses by an amount computed with Eqs. (5) and (6), as the difference between Bel and Pl. In this way we take into account the imprecision

on the estimated sigmoid, and the smaller the distance of a sample to the SVM hyperplane, the higher the amount of imprecision that will be considered. In our example, sample x_1 stands in the uncertain area between support vectors ($|s_{x_1} < 1|$), so that we know that a small change in the logistic optimal parameter estimation could possibly lead to a significant change in the probability estimate. On the contrary, sample x_2 has an associated SVM score which is relatively high, and thus represents a test sample for which the classification is quite sure and will not easily change even in presence of calibration inaccuracy. With the proposed bba allocation in the context of BF framework, we are therefore able to assign a higher value of imprecision to sample x_1 with respect to x_2.

2.3. Bba definition based on pixel neighborhood information

Regarding the second type of imprecision, namely the spatial one, it comes from the fact that in the context of high-density crowd pedestrian detection strong occlusions make the head of each pedestrian barely visible. Besides, due to the specific geometry of the recordings, each head corresponds to few pixels. The most effective head detectors are based on features computed in sub-windows around the pixel of interest, which further increases the spatial imprecision of the detection. For this reason, following the preliminary work we introduced in [49], we model the spatial imprecision due to the close resolutions of object (head) and descriptor respectively by performing opening operation in the spatial domain to discount the bba taking into account the neighborhood heterogeneousness.

In particular, the bba allocation proposed in [39] is able to take into account both types of imprecision, aiming to be more robust to possible imperfections of the learned sigmoid from which the mapping from SVM scores to probability values is made, while at the same time taking into account the information coming from neighboring pixels in the image space. Practically, we process two successive discounting steps on the initial Bayesian bba derived from the learned sigmoid. Firstly, having learned the sigmoid of classifier i by logistic regression, we define bbas to model the imprecision due to possible errors in the calibration, by applying an erosion operator in the 2D space where SVM calibration scores are projected with respect to their label. Then, we

increase the mass on Θ discounting the previous bba by performing a morphological opening operation, this time in the image space, to take into account neighbor pixels information based on the assumption that they are likely to belong to the same class.

In the following, notations are slightly modified from [39] to let appear the dependency from both the considered pixel or sample (noted x) and the considered classifier (noted i). More in detail, dilation δ_w and erosion $\mathcal{E}_w$ operators depending on the structuring element of width w are composed with the calibrated sigmoid σ_i relative to classifier i, in order to derive the two different sigmoid functions, denoted $(\delta_w \circ \sigma_i)$ and $(\mathcal{E}_w \circ \sigma_i)$, representing Pl and Bel function values evaluated on the score s_x relative to sample x (cf. Fig. 1). This takes into account the imprecision of the calibration step. Then, for each pixel x and classifier i independently, we derive the 'one-time' discounted bba $\widetilde{m}_{x,i}$, so that at the end of this step we get a map (image) of bbas $\{\widetilde{m}_{x,i}, x \in \mathcal{P}\}$, where $\mathcal{P}$ is the pixel domain. This image $\widetilde{\mathcal{M}_i}$ is composed by four layers corresponding to the mass values of any hypothesis in $\left\{\emptyset, H, \overline{H}, \Theta\right\}$, respectively. Then, applying an opening to $\widetilde{\mathcal{M}_i}$ second and third layers (i.e., the ones corresponding to singleton hypotheses), and increasing accordingly the Θ layer values, the map $\mathcal{M}_i$ of the final bbas $m_{x,i}$ is derived. This allows us to model the imprecision of the detectors.

Specifically, with s_x being the SVM score associated to pixel x, we have:

$$\left\{\begin{array}{rcl} \forall x \in \mathcal{P}, \widetilde{m}_{x,i}(H) & = & (\mathcal{E}_w \circ \sigma_i)(s_x), \\ \forall x \in \mathcal{P}, \widetilde{m}_{x,i}(\overline{H}) & = & 1 - (\delta_w \circ \sigma_i)(s_x), \\ \forall x \in \mathcal{P}, \widetilde{m}_{x,i}(\Theta) & = & 1 - \widetilde{m}_{x,i}(H) - \widetilde{m}_{x,i}(\overline{H}). \end{array}\right.$$

where σ_i is the learned sigmoid for classifier i and $(\mathcal{E}_w \circ \sigma_i)$ and $(\delta_w \circ \sigma_i)$ its eroded and dilated results with a (flat) structuring element of width w, applied in the score space. Then, in the image space,

$$\left\{\begin{array}{rcl} \mathcal{M}_i(\emptyset) & = & \{0\}_{x \in \mathcal{P}}, \\ \forall A \in \left\{H, \overline{H}\right\}, \mathcal{M}_i(A) & = & \gamma_a\left(\widetilde{\mathcal{M}_i}(A)\right), \\ \mathcal{M}_i(\Theta) & = & \{1\}_{x \in \mathcal{P}} - \mathcal{M}_i(H) - \mathcal{M}_i(\overline{H}), \end{array}\right.$$

Table 2: Neighborhood spatial arrangement for samples x_1 and x_2. Corresponding mass allocations are reported in Table 3.

neighborhood of sample x_1			neighborhood of sample x_2		
	x_{11}			x_{21}	
x_{14}	$\boldsymbol{x_1}$	x_{12}	x_{24}	$\boldsymbol{x_2}$	x_{22}
	x_{13}			x_{23}	

Table 3: Example of proposed bba allocation after discounting based on SVM scores, for neighborhood of samples x_1 and x_2 spatially arranged as reported in Table 2. Bba allocation for samples x_1 and x_2 is already reported in Table 1.

	$\widetilde{m}_{x_{11}}$	$\widetilde{m}_{x_{12}}$	$\widetilde{m}_{x_{13}}$	$\widetilde{m}_{x_{14}}$	$\widetilde{m}_{x_{21}}$	$\widetilde{m}_{x_{22}}$	$\widetilde{m}_{x_{23}}$	$\widetilde{m}_{x_{24}}$
H	0.8	0.2	0.7	0.01	0.95	0.94	0.98	0.95
$\overline{H}$	0.19	0.4	0.2	0.8	0.04	0.03	0.01	0.03
Θ	0.01	0.4	0.1	0.19	0.01	0.03	0.01	0.02

where $\mathcal{M}_i(A)$ is the layer image associated to hypothesis A, $\forall A \in 2^\Theta$, and γ_a is the opening operator of parameter a applied in the image domain.

As in [49], a spatial Gaussian structuring element fitted in a window of radius a is used, to better take into account the spatial consistency. Note that the two morphological operations described are not commutative, since they are applied in two different spaces, i.e. score and image domains, and we find it more natural to firstly consider the imprecision due to the calibration step and later consider the imprecision in the spatial context.

Let us continue with the toy example proposed in the previous section. Table 2 shows the spatial arrangement of neighbor samples around the considered x_1 and x_2. Let us suppose that neighbors have associated bbas reported in Table 3 after bba allocation based on SVM scores. Bba allocation for samples x_1 and x_2 is already reported in Table 1. Note that the spatial arrangement of the samples is fully independent from their position in the score space. It is evident in the example that x_2 has a more homogeneous neighborhood with respect to x_1. This reflects in a higher discounting for sample x_1 (for simplicity, in the example applying erosion with a flat 4-connectivity

Table 4: Example of bba allocation for samples x_1 and x_2. From the bbas based on imprecise score we derive the final bbas applying a second discounting based on neighboring pixels heterogeneity (in this example, with flat 4-connectivity structuring element).

Sample	**Imprecise score-based bba**	**Final bba**
x_1	$\widetilde{m}_{x_1}(H) = 0.12$	$m_{x_1}(H) = 0.01$
	$\widetilde{m}_{x_1}(\overline{H}) = 0.49$	$m_{x_1}(\overline{H}) = 0.19$
	$\widetilde{m}_{x_1}(\Theta) = 0.39$	$m_{x_1}(\Theta) = 0.8$
x_2	$\widetilde{m}_{x_2}(H) = 0.95$	$m_{x_1}(H) = 0.94$
	$\widetilde{m}_{x_2}(\overline{H}) = 0.01$	$m_{x_1}(\overline{H}) = 0.01$
	$\widetilde{m}_{x_2}(\Theta) = 0.04$	$m_{x_1}(\Theta) = 0.05$

structuring element). Note that with the Bayesian allocation we would have assigned to x_1 a high mass on $\overline{H}$, while taking into account the two types of imprecision we end up with a bba having a high value of ignorance, that will not contribute a lot in the conjunctive combination with the other classifiers. On the contrary, the final bba allocation of x_2 reflects its Bayesian counterpart, since its calibrated score is quite reliable and its neighborhood is homogeneous.

2.4. BBAs combination

Considering the N different descriptors, N bbas are defined as explained for every sample x. According to the bba obtained from descriptor i, the uncertainty of a head presence in the pixel associated to x ranges between $Bel_{x,i}(H) = m_{x,i}(H)$ and $Pl_{x,i}(H) = m_{x,i}(H)+m_{x,i}(\Theta)$, so that $m_{x,i}(\Theta)$ represents the imprecision on the uncertainty value provided by i^{th} descriptor for the given sample. In the proposed model, the uncertainty comes from the binary classifier score, whereas the imprecision comes both from the initial score calibration and from spatial heterogeneity of uncertainty values within the considered structuring element.

Finally, the combination between bbas can be performed. As the descriptors are considered *cognitively* independent, the orthogonal sum or its unnormalized version, the conjunctive combination rule [41], are well-suited for this task. For two sources

m_1 and m_2, the conjunctive combination rule is defined by

$$\forall A \in 2^{\Theta}, m_{1\textcircled{\cap}2}(A) = \sum_{\substack{(B,C)\in 2^{\Theta}\times 2^{\Theta}, \\ B\cap C=A}} m_1(B)\, m_2(C). \tag{7}$$

In our case where $|\Theta| = 2$, and considering $m_{x,i}$ bbas allocation, the analytical result may be easily derived:

$$\begin{cases} m_x(A) &= \sum_{\substack{(B_1,\ldots,B_N)\in\{A,\Theta\}^N, \\ \exists i\in[1,N] s.t. B_i=A}} \prod_{i=1}^{N} m_{x,i}(B_i), \forall A \in \left\{H, \overline{H}\right\}, \\ m_x(\Theta) &= \prod_{i=1}^{N} m_{x,i}(\Theta), \\ m_x(\emptyset) &= 1 - m_x(H) - m_x\left(\overline{H}\right) - m_x(\Theta). \end{cases}$$

The result is thus a single four-layer map $\mathcal{M}$ of bbas m_x, where the overall ignorance is reduced as a result of the combination, but at the same time a conflict component may appear in each pixel.

Finally, for every sample, the decision is taken from its corresponding m_x. Several rules have been proposed in the literature. Most popular ones only consider singleton hypotheses (in order to avoid ambiguous decision) and are based on functions that have a probabilistic interpretation: maximum of plausibility, credibility, or pignistic probability [41].

Pignistic probability in particular can be used to give a probabilistic interpretation to the bbas. Since in our setting $|\Theta| = 2, \forall A \in \Theta$

$$BetP_x(A) = \frac{1}{1 - m_x(\emptyset)} \cdot \left(m_x(A) + \frac{m_x(\Theta)}{2} \right). \tag{8}$$

This allows us to assign a probabilistic interpretation to the resulting bba associated to each sample, so that we will obtain a single-layer $BetP(H)$ image map where at every pixel the $BetP_x(H)$ value will be differently normalized on the basis of the conflict value included in m_x, represented by the mass on the empty set. Generally, from the $BetP(H)$ map, statistics for quantitative evaluation of the detection results are derived. Besides doing this, in our study, obtained bbas m_x are also taken into account for the selection of new samples during the AL process, as explained in next section.

3. Evidential QBC disagreement measures

In the context of QBC, new samples for the learning process are chosen based on the analysis of the responses of a set or committee of classifiers. Traditionally, generic ensemble learning algorithms are used to build the committee (i.e., bagging or boosting), or a set of weak classifiers are derived from a model changing its intrinsic parameters. In the context of SVM-based learning, among some descriptors which are widely used in pedestrian detection, those which are the best suited in high-density crowds have been recently highlighted in our previous works [50, 49], so that we find it natural to build a set of classifiers with them.

After having built $\mathcal{C}$, the committee of classifiers of cardinality $|\mathcal{C}| = N$ sources, QBC relies on some heuristics to measure the disagreement among them, in order to find the most informative samples to add to the training set $\mathcal{L}$.

In the following, we investigate traditional disagreement metrics such as Soft Vote Entropy and KL divergence, as well as new evidential-based disagreement measures, with diversity among samples ensured by a minimum Euclidean distance applied in the spatial domain between instances already in the training set and in the current batch.

3.1. Traditional disagreement measures in QBC and their limitations

Specifically, given the set of mutually exclusive hypotheses $\Theta = \left\{H, \overline{H}\right\}$ and $\mathcal{C}$ the committee of classifiers of cardinality N, Soft Vote Entropy asks for the label of the unlabeled sample such that:

$$x^*_{SVE} = \underset{x \in \mathcal{U}}{\operatorname{argmax}} \sum_{y \in \Theta} P_{\mathcal{C}}\left(y \mid x\right) \log \left(\frac{1}{P_{\mathcal{C}}\left(y \mid x\right)} \right), \tag{9}$$

where $\mathcal{U}$ is the set of unlabeled samples ($\mathcal{U} \subset \mathcal{P}$), and $P_{\mathcal{C}}\left(y \mid x\right) = \frac{1}{N} \sum_{i=1}^{N} P_i\left(y \mid x\right)$ is the average or *consensus* probability that y is the correct label according to the committee. Soft Vote Entropy is thus essentially an ensemble generalization of entropy-based uncertainty sampling. The $\log$ function, here and from now on, represents the logarithm to the base 2.

On the other hand, the KL divergence strategy adds samples to the training set such that:

$$x^*_{KL} = \underset{x \in \mathcal{U}}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{D}_{KL} \left(P_{x,i} \| P_{x,\mathcal{C}} \right), \tag{10}$$

where $P_{x,i} = P_i \left(y \mid x \right)$ and $P_{x,\mathcal{C}} = P_{\mathcal{C}} \left(y \mid x \right)$ for simplicity of notation, while $\mathcal{D}_{KL}$ is the KL divergence that quantifies the disagreement as the average divergence between the prediction of each classifier i in the committee and the consensus $P_{\mathcal{C}}$, and is defined by

$$\mathcal{D}_{KL} \left(P_{x,i} \| P_{x,\mathcal{C}} \right) = \sum_{y \in \Theta} P_i \left(y \mid x \right) \log \left(\frac{P_i \left(y \mid x \right)}{P_{\mathcal{C}} \left(y \mid x \right)} \right). \tag{11}$$

The conceptual difference behind SVE and KL resides in the way they quantify the *disagreement*. Considering a committee of classifiers, the consensus probability $P_{\mathcal{C}} \left(y \mid x \right)$ between them could be uniform in two different cases. Firstly, all the classifiers have an uniform distribution among the hypotheses, so that the consensus distribution is also uniform. Secondly, the classifiers strongly disagree between them, but since the consensus is an average between their responses, it ends up being uniform among all the hypotheses as well. In the first case, all the classifiers agree that the label is uncertain, while in the second case they strongly support a different label. Since SVE only considers consensus, it cannot distinguish between the two cases. On the other hand, KL divergence would favor only samples with uncertain consensus because of conflicting predictions given by the classifiers.

Besides these highlighted limitations, the mentioned measures do not exploit the possible information arising from the combination among the committee members, and the final result on which evaluation is performed is not taken into account in the selection of the new samples.

3.2. Proposed evidential disagreement measures

On our side, after having performed the combination between the various sources in the BF framework, the result is the map $\mathcal{M}$ where at each pixel x of the image corresponds a bba m_x that incorporates a different evidence of belonging to a certain

Table 5: Evidential entropy definitions given bba m with discernment frame Θ

Reference	**Entropy formulation**
Höhle [51]	$H_O(m) = \sum_{A \in 2^\Theta} m(A) \log\left(\frac{1}{Bel(A)}\right)$
Yager [52]	$H_Y(m) = \sum_{A \in 2^\Theta} m(A) \log\left(\frac{1}{Pl(A)}\right)$
Nguyen [53]	$H_N(m) = \sum_{A \in 2^\Theta} m(A) \log\left(\frac{1}{m(A)}\right)$
Pal et al. [54, 55]	$H_P(m) = \sum_{A \in 2^\Theta} m(A) \log\left(\frac{\lvert A\rvert}{m(A)}\right)$
Dubois and Prade [56]	$H_{DP}(m) = \sum_{A \in 2^\Theta} m(A) \log\left(\lvert A\rvert\right)$
Lamata and Moral [57]	$H_{LM}(m) = H_Y(m) + H_{DP}(m)$
Deng [58]	$H_D(m) = H_N(m) + \sum_{A \in 2^\Theta} m(A) \log\left(2^{\lvert A\rvert} - 1\right)$
Jiroušek and Shenoy[59]	$H_{JS}(m) = \sum_{x \in \Theta} Pl\_P(x) \log\left(\frac{1}{Pl\_P(x)}\right) + H_{DP}(m)$
Jousselme et al. [60]	$H_J(m) = \sum_{x \in \Theta} BetP(x) \log\left(\frac{1}{BetP(x)}\right)$

class (i.e., head or not head), as well as a component of ignorance that remains after the combination, and conflict between the sources, i.e. the masses on Θ and $\emptyset$ respectively that come from the conjunctive combination. We can therefore extend the concept of Soft Vote Entropy to the evidential framework, to define new evidential measures of disagreement among committee members. The Maximum Entropy (ME) strategy will add to the training set sample such that:

$$x^*_{ME} = \underset{x \in \mathcal{U}}{\operatorname{argmax}}\, H(m_x), \tag{12}$$

where in our case m_x is the bba associated to the unlabeled sample x, obtained after the explained bba allocations and conjunctive combination, and $H(\cdot)$ is a definition of the entropy function in the evidential domain.

Several definitions of *evidential entropy* have been proposed over the past decades, with the aim of measuring the degree of total uncertainty of a bba, but a formulation satisfying all the desired properties still remains an open issue.

Table 5 summarizes some popular definitions, that we intend to investigate as functions to select the new training points. Some of them, like Höhle [51], Yager [52] and Nguyen [53] definitions are only able to measure the conflict portion of uncertainty. Pal definition [54, 55] is an extension of Nguyen's one, taking into account also the

cardinality of each focal element. The definition given by Dubois and Prade [56], on the contrary, captures only the non-specificity portion of uncertainty, quantifying how a bba is imprecise. The most non-specific bba is given by the categorical bba having $m(\Theta) = 1$, while the most specific bbas are the Bayesian ones, so that non-specificity is a measure of how a bba is fragmented among the various hypotheses. The formulation given by Lamata and Moral [57] and the more recent Deng [58] and Jiroušek and Shenoy [59] ones, combine both conflicting and non-specificity components in different ways. Regarding the conflicting part, Lamata et al. uses Yager's definition which relies on the plausibility function, Deng uses Nguyen's formulation while Jiroušek et al. interprets it in a completely different way, as the Shannon's entropy of the plausibility probability function Pl_P [61], an alternative method to pignistic transformation for translating bbas into probabilistic framework. Regarding the non-specificity component, Lamata et al. and Jiroušek et al. rely on Dubois and Prade definition, while Deng provides a brand new formulation. Alternatively, Jousselme et al. [60] firstly perform a pignistic transformation from bba to probability mass function through $BetP$, and then apply Shannon's entropy on it. A similar definition, called pignistic entropy, appears in [62], in the context of the Dezert-Smarandache Theory (DSmT) [63, 64], that is a variant of the classical Dempster-Shafer Theory (DST). Since we indeed rely on DST, we refer in the following to Jousselme's definition. The advantage of such a formulation for our application is that since it is based on the $BetP$ function, there is a direct link between it and the final map we use for decision and, possibly, crowd density evaluation [50].

Besides entropy-based criteria, the masses on Θ and $\emptyset$ can be directly exploited as indicators for the selection of the new samples. It is possible to directly derive two simple strategies, based on Maximum Ignorance (MI) and Maximum Conflict (MC):

$$x^*_{MI} = \underset{x \in \mathcal{U}}{\operatorname{argmax}}\, m_x(\Theta), \tag{13}$$

$$x^*_{MC} = \underset{x \in \mathcal{U}}{\operatorname{argmax}}\, m_x(\emptyset). \tag{14}$$

where in our case m_x is the bba associated to the unlabeled sample x, obtained after the explained bba allocations and conjunctive combination.

Equation (13) favors the selection of new points for which all the classifiers do not have enough information to assign them to one of the classes, i.e. samples with maximal mass on the compound hypothesis. On the contrary, Eq. (14) supports the selection of points on which the classifiers disagree the most about their actual label, i.e. samples with maximal mass on the empty set. In Eq. (14) we choose to use a measure of total conflict derived from the conjunctive combination rule as disagreement measure. In [65, 66] total conflict is separated into internal and external components. Internal conflict quantifies the (self-)inconsistency of the i^{th} source, while external conflict is only based on the interaction between sources and does not integrate any self-inconsistency. The authors of [67] in particular agree with this subdivision, and they propose conflict measurements based on contour functions, making no a-priori assumptions regarding the dependence between sources.

The concepts of conflict and ignorance have already been used in the context of single classifier uncertainty sampling-based AL in [68], but with different meanings from those in the BF framework. In their work, conflict models the extent to which a new query point lies in the conflict region between two or more classes (whereas for us it refers to conflicting beliefs from different classifiers), while ignorance represents the distance of a new query point from the training samples seen so far, so that it is higher in areas of the version space not represented yet (while for us it is higher when for all the classifiers the point resides in their uncertainty area - in a sense, the two definitions are completely the opposite). Always in the different context of uncertainty sampling, in [69] there is a distinction between insufficient-evidence and conflicting-evidence uncertainties, but the concept of *evidence* does not refer to BF framework, but it is rather measured as a weighted similarity of a given sample to the support vectors.

We expect that the inclusion of samples with high ignorance or conflict will be beneficial for the learning process, respectively in order to sharpen the decision boundaries between the classes for all the classifiers and to reduce overall conflict between the various sources. However, the former strategy exploits examples which are near the current decision margins in all the feature spaces, and it is not able to solve possible conflicts but it just adjusts the boundaries, while the latter allows for an exploration of the version spaces to select points which are not yet represented by the current models,

Table 6: Example of evidential-based disagreement measures

	$\emptyset$	H	$\overline{H}$	Θ	$BetP(H)$	H_O	H_Y	H_N	H_P	H_{DP}	H_{LM}	H_D	H_{RP}	H_J
m_{x_1}	0.01	0.1	0.1	**0.79**	0.5	0.67	0.03	0.93	1.72	**0.8**	0.83	2.19	**1.8**	**1**
m_{x_2}	**0.79**	0.1	0.1	0.01	0.5	**1.02**	**0.88**	1.23	1.27	0.05	**0.94**	1.3	1.05	**1**
m_{x_3}	0.4	0.1	0.1	0.4	0.5	0.86	0.09	**1.25**	**1.92**	0.66	0.75	**2.31**	1.66	**1**
m_{x_4}	0.1	0.79	0.01	0.1	0.93	0.24	0.05	0.59	0.7	0.11	0.16	0.76	0.61	0.35

but it could be prone to select outliers. In this sense they are complementary strategies, and they should be used in conjunction with a criterion able to balance them. Alternatively, we expect entropy-based disagreement to be able to naturally find a trade-off between them as a measure of information gain.

Table 6 shows an example of four bbas associated to different samples, and the decision about which sample to query based on the different evidential criteria (i.e. the sample related to the bold value in each column). In particular, m_{x_1} has a high component of ignorance, m_{x_2} is a very conflicting bba, m_{x_3} is not committed and at the same time has a high amount of both ignorance and conflict, while m_{x_4} is committed about H hypothesis. The value of $BetP(H)$ is also shown, to highlight the fact that the probabilistic framework assigns the same value to the first three bbas even if they are intrinsically very different one from the others. As we expect, no measure selects m_{x_4} to be added to the training set, since it is quite committed and it would not provide much information. On the contrary, the first three bbas are selected based on the different measures. A clear limitation of MI and MC criteria is that they fail detecting bbas with relatively high values of both conflict and ignorance: MI selects m_{x_1} while MC selects m_{x_2}, but they do not consider m_{x_3} at all, even if it represents a potentially interesting sample to add to the training set. Conversely, entropy-based criteria are able to better consider the relative repartition of masses through the various hypotheses. Using Höhle and Yager definitions of entropy, m_{x_2} is selected, highlighting their tendency to detect conflicting instances. Nguyen and Pal favor the selection of m_{x_3}, prioritizing samples which are both not very committed and conflicting, even if Nguyen is more sensitive to conflict while Pal gives more importance to the ignorance component. Dubois and Prade's formulation of entropy favors samples with high ignorance, not being able to capture the conflict component. Among the three composite formulation that aim at taking into account both conflict and non-specificity (i.e., Lamata

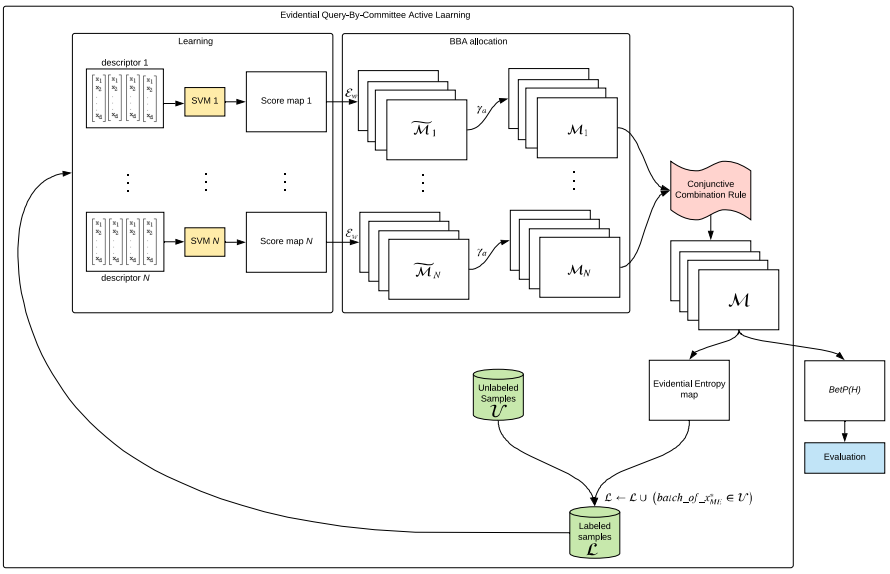


Figure 2: Evidential Query-By-Committee Active Learning flowchart.

and Moral, Deng, Jiroušek and Shenoy), we can notice that they all prioritize different samples, but there is only a slightly difference among the entropy values associated to the first three bbas. This suggests the fact that they would probably select the three of them to be part of the same batch. In the same way, Jousselme's definition based on $BetP(H)$ encourages a diversity in terms of bbas in the same batch, allowing to tackle different types of issues at the same time (i.e., conflicting and/or not committed bbas).

3.3. Global overview of the proposed evidential QBC process: from bba allocation to new sample selection

Figure 2 shows the complete flowchart of the proposed evidential QBC method. After the traditional learning step, BF framework is involved in three important operations, namely in the bba allocation procedure through successive discounting, in the combination of sources that allows us to obtain a $BetP(H)$ map used for evaluation, and in the derivation of evidential entropy map which guides the selection of the most informative samples to add to the training set for the subsequent iteration of the active

learning procedure.

The proposed evidential QBC differs from the traditional one. First of all, from the score maps given by SVM classifications we do not derive probabilistic maps through logistic regression, but we perform a bba allocation that takes into account two possible sources of imprecision, namely in the estimation of the sigmoid parameters to perform logistic regression and, later, in the image space. Then, the conjunctive combination rule is able to take into account the information provided by the different sources, discounted pixel-wise accordingly to their evaluated reliability. At this stage, the obtained bba map $\mathcal{M}$ can be used either for evaluation, through the computation of the $BetP(H)$ map, or to compute the evidential entropy map, from which the samples with maximum entropy are extracted and added to the labeled samples set $\mathcal{L}$. Note that in case of Maximum Ignorance or Maximum Conflict criteria, the evidential entropy map would not be computed, and the samples would directly be chosen maximizing ignorance and conflict channels, $\mathcal{M}(\Theta)$ and $\mathcal{M}(\emptyset)$ respectively.

In the following section, we will investigate all the proposed evidential-based disagreement measures as well as the traditional ones in the context of our application.

4. Experiments

4.1. Data and features used

4.1.1. Dataset

We tested our proposed fusion method on high-density crowd images acquired at Makkah during Hajj [70]. The camera we used is a robotic camera (AVT Guppy PRO) mounted statically in order to observe the high-density pilgrim crowd, and providing gray-level regular images (visible spectrum). The camera provides thus a video sequence of the crowd (at a frame-rate of 8Hz). For the training, calibration and evaluation of the head detectors, we use images extracted at distant moments (in order to establish a full level of independence among the images used). Each image instance contains in the analyzed Region of Interest (corresponding roughly to the lower half of the scene) a high number of objects to detect (about 900-1000 heads) due to the high density.

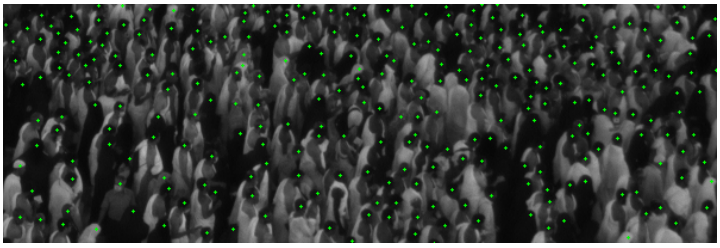


Figure 3: Patch with ground-truth dotted annotation

Figure 3 shows a patch from an image of the dataset, highlighting the difficulty of the problem since the heads are barely visible and many occlusions occur. We performed a dotted annotation in the head centers for the training images, such that the ground-truth so obtained can be used as *oracle* to assign the correct label to the samples selected for querying by AL. Even though in Makkah the crowd follows a general direction, there is a significant degree of head appearance variability due to gender, type of head cover, and most importantly, to the various degrees of occlusion coupled with the small size of the targets. For annotating a single image (clicking on the heads exhaustively), a human annotator requires typically half a day of work, and approximately 20% of the heads are so difficult to annotate that the human needs to look in the previous and the next frames in order to take a head/not head decision (something which our algorithm cannot do, as it performs the detection only in the current frame). As it is possible to see from the image, another problem in this type of scenes is the high data imbalance between positive and negative samples (i.e., pixels belonging or not to a head, respectively), stressing the importance of finding an effective strategy to select significant samples.

4.1.2. Pedestrian detectors in dense crowds

In the difficult context of high-density crowds, simple detectors relying on appearance cues, such as local color histogram which may be associated to skin, hair or clothes are not well suited, since the object resolution needs to be relatively high and the color spaces may not be discriminative enough. Even worse, many surveillance cameras provide only gray level data, as in our case, so that it is not possible at all

to exploit color information. In the same way, common face detectors such as Viola-Jones [71] are unsuited, since pedestrian faces are not detailed enough. Among some descriptors which are widely used in SVM-based pedestrian detection, those which are the best suited in the context of high-density crowds have been recently highlighted in [49].

Related to the image gradient, the Histogram of Oriented Gradients (HOG) descriptor [10] is very popular and has exhibited in various contexts excellent performances when used in conjunction with a histogram intersection kernel (HIK) [9]. The contour related to the specific shape of the head and shoulders is indeed highly discriminative, but it may fade away due to clutter. For this reason, it is important to consider as well some descriptors aimed at other characteristics than shape.

To this extent, some descriptors related to texture representation are the Local Binary Pattern (LBP) operator [11] and Gabor filter banks [72]. The former is traditionally employed in texture classification, and it has been successfully used in pedestrian detection due to its reasonable robustness to occlusion provided by its local sampling strategy, while the latter have been used for head detection [72] to encode the local frequencies and orientations. Regarding LBP, we rely for learning on a χ^2 kernel function which has been shown to be positive definite and suited for data generated from histograms [73], while for the Gabor-based descriptor we consider a RBF kernel.

Besides these popular pedestrian detectors, the DAISY [74] descriptor, usually used in the field of stereo matching, has been successfully employed for the first time for head detection in these difficult crowd scenes in [49], together with HIK. Its Gaussian smoothing, along with the sampling overlap, naturally enforces spatial consistency.

4.2. Evidential QBC result analysis

4.2.1. Comparison between the proposed evidential disagreement measures

For the QBC algorithm, we thus build the committee $\mathcal{C}$ of classifiers with the four cited SVM pedestrian detectors, namely HOG, LBP, Gabor and DAISY. Such a committee is quite heterogeneous since each classifier contributes providing a different view of the data, so that the explained fusion strategy is applied at every iteration, both to obtain the image map of the $BetP(H)$ on which we compute statistics, and to

choose the samples to add to the training set on the basis of the different evidential-based proposed heuristics.

In the context of AL, the choice of the evaluation metrics is not trivial. The recent study carried out by [75] indeed have pointed out that most of the evaluations of AL approaches in the literature have focused on a single performance measure, and have shown that the improvements provided by AL for one performance measure often comes at the expense of another measure. Besides this, the most used metric is accuracy, which intrinsically depends on the choice of a threshold so that a question arises about how much of the observed improvement is due to the effective learning and how much of it is simply due to a shift in the optimal decision threshold. Moreover, accuracy metric is not relevant in presence of highly imbalanced data. To solve this last problem, popular measures are Precision, Recall, and F1 score, but they still require a threshold. For all these reasons, we choose to evaluate our method on the basis of two different measures, which do not depend on a threshold and at the same time are suited for imbalanced data, namely Area Under Precision-Recall Curve (AUPRC) and Precision-Recall Break Even Point (PRBEP) (i.e., the value of the curve where Precision is equal to Recall). These two metrics are computed on the $BetP(H)$ map, applying non maxima suppression (NMS) at every threshold to identify the targets, setting the radius of a head to $r = 3$, with $2r + 1$ minimum distance between two maxima (head centers) in order to avoid overlapping detections.

We conducted our tests starting from a random training set of 500 samples arriving to 2000 samples, with a batch size of 100 samples per iteration added on the basis of the discussed disagreement measures. Figure 4 shows the AUPRC and PRBEP for every iteration using the proposed Maximum Entropy (ME) with the different evidential entropy definitions, Maximum Conflict (MC) and Maximum Ignorance (MI) criteria. It is possible to see an improvement of both metrics with all the investigated disagreement measures, stressing the robustness of the method and the fact that the approach is well-suited to our application. All the curves tend to flatten towards the end of the process, which means that the final number of samples represents a suitable training set size. An interesting consideration is that some curves have higher performance even when relying on a small size of the training set (i.e., are faster to converge). This

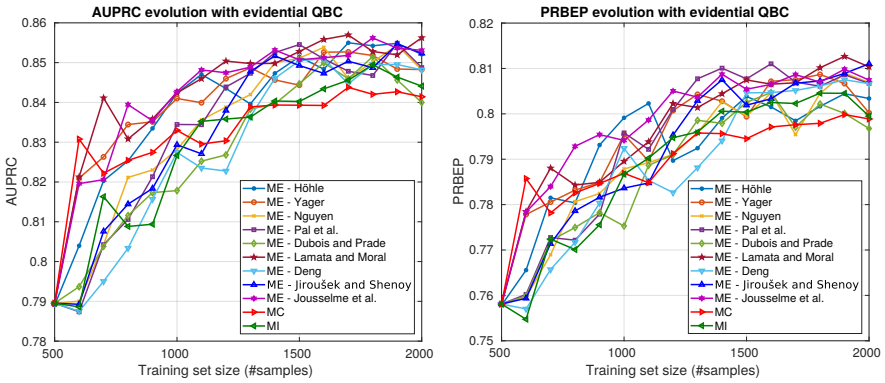


Figure 4: AUPRC and PRBEP at every iteration using ME criterion with different evidential entropy disagreement measures, MC and MI criteria.

means that those query strategies are immediately able to select the most informative samples to add to the training set. There are indeed some differences among the results achieved using the various definitions. It is clear that entropy formulations which focus on conflict (e.g., the Yager one) provide better results with respect to Dubois-Prade definition which focus only on the non-specificity portion of uncertainty, already in presence of a small training set size. Moreover, considering both imprecision and conflict components seems to be beneficial, in particular using Lamata and Moral's composite definition. Note that also the the simpler Jousselme's entropy-based criterion appears quite beneficial both in terms of AUPRC and PRBEP. In general, the best strategies appear to be the ones that encourage *diverse* samples inside the same batch in terms of bba structure, that is to say, both conflict and ignorance components have to be taken into account, with a slight preference for samples with conflicting bbas.

Considering the results obtained with the two simple evidential criteria based solely on conflict and ignorance indeed, these approaches do not reach the performance of entropy-based disagreements. As expected, selecting the samples on the basis of maximum conflict allows for a steeper improvement at the beginning, where exploration of the version space is very important, but after some iterations the curves tend to flatten. On the contrary, the samples with high values of ignorance are mostly useful when the size of the training set begins to be consistent, and it becomes important to exploit the

current feature spaces to adjust the boundaries. This behavior reflects the importance to pass from an initial exploration to a final exploitation of the data. To this extent, evidential QBC based on Maximum Entropy criterion is able to naturally find a trade-off between the two.

In the following, we choose Lamata and Moral's entropy definition as the more competitive criterion among the evidential entropy formulations. Indeed, it outperforms other formulations when considering AUPRC metric, which is a key indicator since it takes into account the whole Precision-Recall curve, and at the same time has good performance in terms of PRBEP.

4.2.2. Comparison with traditional approaches

In order to evaluate the benefit for the active learning procedure of the proposed bba allocation used in conjunction with evidential disagreement measures, Fig. 5 reports the curves related to two different levels of comparison.

Firstly, we evaluate the difference compared to a result reached with purely probabilistic reasoning. We perform the Bayesian bba allocation from the output of each classifier after Platt's regression, without applying any discounting neither in the score space nor in the image space, and we apply the normalized conjunctive combination rule (i.e. Dempster rule): in this way, the classifiers combination boils down to simple product of probabilities. Then, on the resulting probabilistic map, we apply the traditional SVE and KL disagreement measures, as well as a baseline that simply adds randomly drawn samples at every iteration. Moreover, to quantify exactly the benefit of the proposed bba allocation over the Bayesian one, we aim at converting the proposed evidential disagreement measures to the Bayesian framework. MC and MI do not apply, since Bayesian bbas have null masses on conflict and ignorance respectively. Transposing the evidential entropy definitions to the Bayesian framework, however, we notice that all the formulations (except Dubois-Prade's one which is always null being mostly related to ignorance, and Jiroušek and Shenoy's one) boil down to: $H(m) = m(H)\log\left(\frac{1}{m(H)}\right) + m(\overline{H})\log\left(\frac{1}{m(\overline{H})}\right)$. Clearly, our evidential approach outperforms all the probabilistic ones with respect to both AUPRC and PRBEP. Besides, the fact that there is a consistent gap between the proposed evidential Maximum

Entropy and the corresponding curve in the Bayesian framework (Bayesian ME) indicates that the detector combination with the proposed bba allocation is significantly superior to a simple product of probabilities.

Now, in order to show that the performance gain is not only due to the relevant bba allocation, but also to the good choice of disagreement measure for active learning, we perform the proposed evidential bba allocation, obtaining a $BetP(H)$ map that we interpret as a probability map to compute SVE, KL and the random baseline. This allows us to focus on the benefit of the BF framework vs. probabilistic one only with respect to the new sample selection step, to see exactly the impact of evidential measures in the selection of the new samples being not biased by the detector combination result. The related curves are referred in Fig. 5 as "Semi-evidential", since the BF framework is only involved in the bba allocation and combination but not in the sample selection.

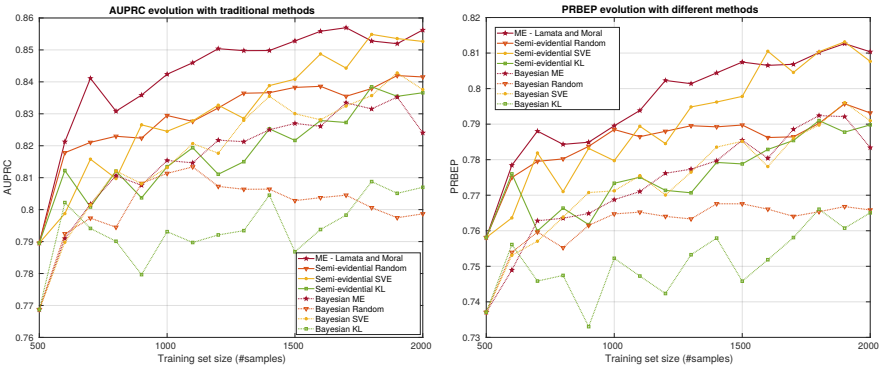


Figure 5: AUPRC and PRBEP at every iteration. Comparison of evidential-based disagreement measures with traditional ones.

Entropy-based criteria, namely SVE and the proposed ME using Lamata and Moral evidential entropy definition, outperform the others, both in terms of AUPRC and PRBEP. However, although reaching almost the same performance as the evidential ME at the very end of the process, SVE is not able to select the most informative samples from the beginning. In particular, entropy-based evidential criterion results to be the best one, due to the ability of BF framework to model in a finer way the actual information contained in each sample, highlighting the importance of the coupling be-

tween the fusion of the classifiers and the definition of the disagreement measures. We can notice how KL strategy, which in its intention should select conflicting samples based on the consensus probability, does not seem to be very efficient in this context, performing even worse than random sampling. This is against what we observed in the comparison of the various evidential-based entropies, where the definitions that focus on the conflict are indeed the most successful ones. This fact shows that the evidential framework is more able to model the conflict among the various committee members, through the mass on $\emptyset$, with respect to the probabilistic framework that models it in terms of divergence from the consensus probability.

4.2.3. Correlation analysis

The aim of correlation analysis between the various disagreement measures is to understand better how they may differ from one another and the similarity between them. To this extent, we apply the proposed MC, MI and ME criteria with all the investigated entropy measures on the basis of the $BetP(H)$ map obtained after bba allocation and combination. Traditional SVE, KL and the random sampling are still performed on the basis of the $BetP(H)$ map obtained after bba allocation, interpreting it as a probability map, to focus only on the new sample selection step.

Figure 6 shows the correlation matrix in terms of percentage of common samples between the different points selected at every iteration on the basis of the investigated criteria, excluding the initial common 500 samples, so that only the ones selected with respect to the various strategies are taken into account in the computation. We do not plot all the iterations but we focus on the first iterations, where variations are more visible, and on the last one in order to give a sight of the general behavior. We note that in general, going on with the iterations, the different training sets tend to diverge, sign that the size of the considered pool of unlabeled samples $\mathcal{U}$ is indeed appropriate in the sense that the various methods have enough freedom being not constrained by the data. Correlation is especially evident considering the various evidential disagreement measures. Many definitions are related to ignorance, and as expected, Dubois and Prade's entropy is very close to it. Yager's entropy and Lamata and Moral's one, on the contrary, are very related one to each other and have a consistent overlap with

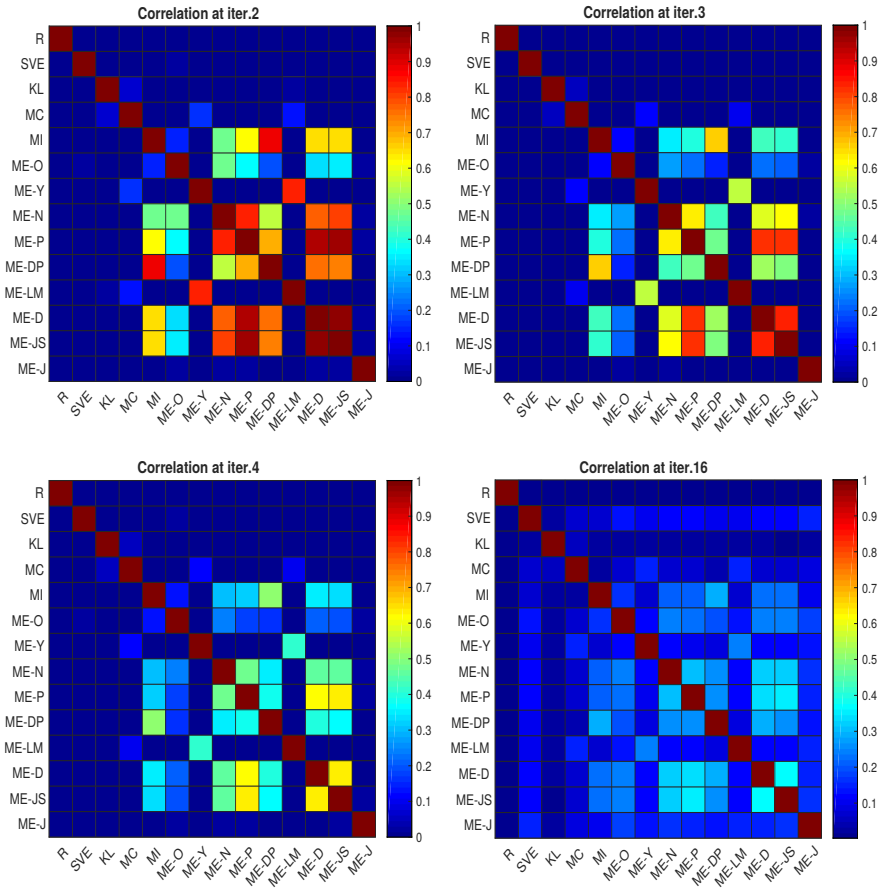
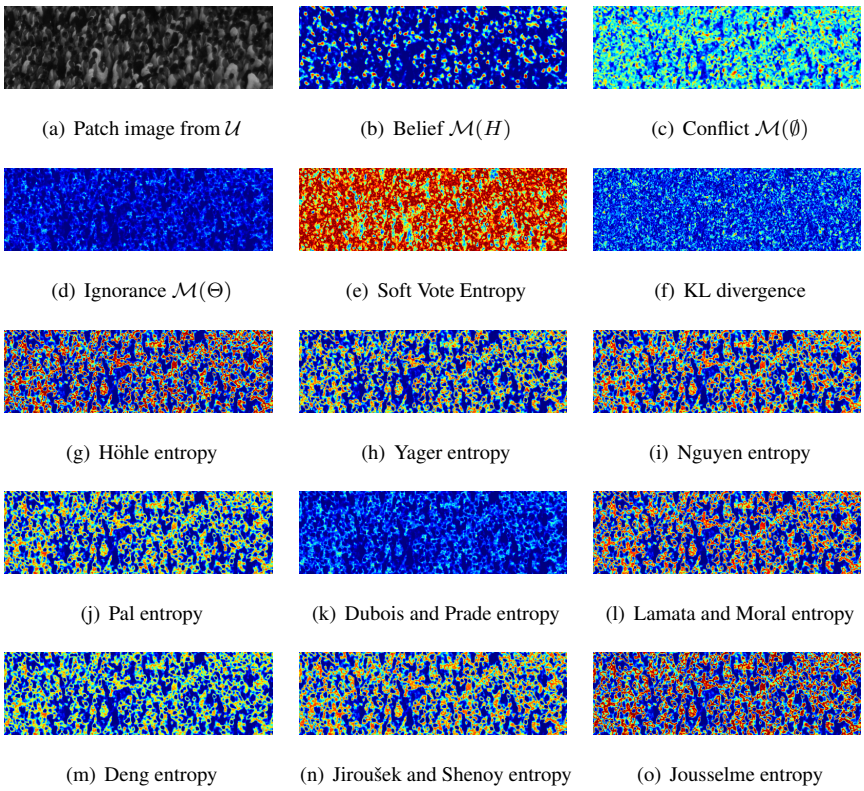


Figure 6: Correlation between samples added during successive AL iterations with different strategies, for the initial iterations and the last one. R = Random, SVE = Soft Vote Entropy, KL = Kullback-Leibler divergence, MC = Maximum Conflict, MI = Maximum Ignorance, ME = Maximum Entropy: O = Höhle, Y = Yager, N = Nguyen, P = Pal et al., DP = Dubois and Prade, LM = Lamata and Moral, D = Deng, JS = Jiroušek and Shenoy, J = Jousselme.



(a) Patch image from $\mathcal{U}$ (b) Belief $\mathcal{M}(H)$ (c) Conflict $\mathcal{M}(\emptyset)$

(d) Ignorance $\mathcal{M}(\Theta)$ (e) Soft Vote Entropy (f) KL divergence

(g) Höhle entropy (h) Yager entropy (i) Nguyen entropy

(j) Pal entropy (k) Dubois and Prade entropy (l) Lamata and Moral entropy

(m) Deng entropy (n) Jiroušek and Shenoy entropy (o) Jousselme entropy

Figure 7: Different maps obtained using the investigated evidential disagreement measures for a selected patch of the unlabeled samples pool $\mathcal{U}$ (in Fig. (a)) with corresponding bba allocation $\mathcal{M}$. SVE and KL maps are shown as well for comparison.

the conflict measure. Nguyen and Pal correlation is also highlighted, and it is easily explainable by the fact that Pal's formulation extends Nguyen's one, taking into account also the cardinality of the focal elements (in our case, in presence of two singleton hypotheses, only the term that refers to the compound set slightly changes). Again, Pal's training set seems very correlated to Jiroušek-Shenoy's and Deng's ones, which are two composite formulations aiming to take into account both conflict and non-specificity. KL divergence seems totally unrelated to any other measure, except for the conflict with a marginal degree.

To better understand the degree of correlation between the different measures,

Fig. 7 shows a visual comparison of the maps obtained with the various entropy definitions for the first iteration of the AL process, so that we can compare them on the basis of the same training set. Figure 7(a) represents a selected part of the unlabeled samples pool $\mathcal{U}$. After the evidential combination of the classifiers, the result is the image map $\mathcal{M}$ of bbas m_x associated to every pixel x, shown in terms of belief in Fig. 7(b), conflict in Fig. 7(c), and ignorance in Fig. 7(d). Soft Vote Entropy 7(e) and KL divergence 7(f) maps are shown as well for comparison with all the investigated evidential entropy definitions. Once again, we notice the correlation between ignorance and Dubois-Prade entropy in Fig. 7(k), while the other entropy definitions seem more correlated to the conflict, although to different extents. In general, evidential entropy maps are able to model in a finer way the actual information contained at every pixel locations, so that the regions of interest for the AL process are better enhanced with respect to SVE and KL.

The figure visually shows where and how the various entropy definitions correlate. While previous Fig. 6 provides only a global estimation of the correlation (scalar value), Fig. 7 allows for a qualitative visualization of the spatial variation of the correlation. Entropy is higher where the individual detectors are discordant, and the images show that this happens frequently on the border of the heads, because the various classifiers provide different detection sizes (e.g. HOG and Gabor provide more localized detections with respect to LBP and Daisy that provide coarser blobs). There are some areas that correspond to a head where entropy is high, and it means that just a part of the classifier committee succeeds in detecting it. We also note that some shoulders of the people may present high entropy values. Specifically, this happens when one or some classifiers miss-classify shoulders as heads due to their similar rounded visual appearance. Finally, it is interesting to visualize that the maps usually agree on the *location* of maximum entropy (borders of the heads, heads detected only by some classifiers, shoulders areas which confuse some classifiers), while at the same time they provide different *amounts* of entropy for the same location, and this is what allows the AL to choose different samples and thus to obtain such diverse training set at the end of the process.

4.2.4. *Global benefit of evidential QBC active learning*

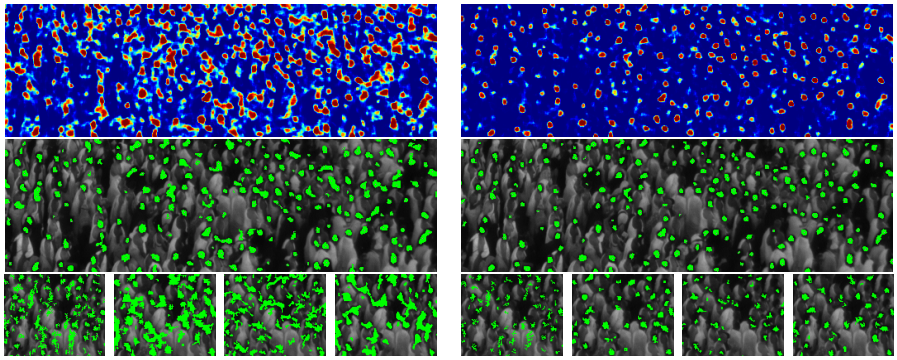


Figure 8: Visual comparison of the detections obtained at the first iteration of the process (500 training samples), on the left, and the last iteration (2000 training samples selected using Lamata and Moral Maximum Entropy criterion), on the right. Results are shown in terms of colormap of the $BetP(H)$ map in the first row, detections at PRBEP in the second row, and the different sources used in the combination in the third row (namely HOG, LBP, Gabor, Daisy).

Figure 8 provides a visual comparison between the first and the last iterations of the process, during which the training set increased from 500 samples (on the left) to 2000 training samples (on the right), selected with the Maximum Entropy criterion using Lamata and Moral's definition. The classification results are shown both in terms of colormap of the $BetP(H)$ in the first row, and detections at the PRBEP threshold in the second row. Moreover, detections using the single sources that compose the committee of classifiers are shown in the last row (HOG, LBP, Gabor, Daisy respectively), in order to highlight their complementarity and the necessity of an adapted fusion between them.While the colormap is useful to identify regions with higher values, and to immediately see that at the end of the process we obtain a less noisy and sharper map, the detections superimposed on the input image are indeed useful to evaluate the actual location of the detections and the presence of false positives (areas with high values which do not correspond to an actual head) or false negatives (heads which are not detected). The detections are provided here for the value of threshold at which precision is equal to recall (i.e. the PRBEP), which is a reasonable compromise since it allows us to have the same number of false positives and false negatives. PRBEP is equal to

0.74 for the first iteration, meaning that at the beginning of the process for this particular threshold 26% of the heads are lost while at the same time 26% of the detections are not actual heads. At the end of the process, PRBEP becomes 0.835, meaning that we obtain an improvement of almost 10% with the proposed approach, both in terms of precision and recall. PRBEP threshold is a traditional operative point for many applications and we find it reasonable to adopt it for visualization purposes. The exact values of the thresholds are $th = 0.8$ for the first iteration (on the left) and $th = 0.55$ for the last iteration (on the right). Although the exact values are not really meaningful in themselves, it is interesting to notice the initial bias toward a high threshold, that can be explained by the fact that at the beginning of the process the training set is balanced (i.e. it has the same number of positive and negative samples), while at the end of the process it tends to have more negative samples, reflecting the actual data distribution.

Overall, the proposed evidential fusion, which is able to take into account imprecision both during calibration and in the image space, results to be suited for this application. Besides, the AL algorithm is able to select samples which are indeed useful to improve the performance of all the classifiers, a fact which results in a significant and visible improvement of the final $BetP(H)$ map. At the end of the AL process, the detections are more localized, sharper and a lot of false positives which were present at the beginning have been successfully removed.

To highlight the importance of coupling the fusion strategy with the AL process, Fig. 9 shows different Precision-Recall (PR) curves, for the various single classifiers as well as for their fusion. The curves are shown for the first and last iterations, in order to illustrate the relative improvement in terms of performance for all the classifiers.

Besides showing that fusion results are better than individual detectors (which is not mandatory true especially in case of poor initial detectors and underlines the importance of the proposed bba allocation), the figure has two main purposes. Firstly, it shows the improvement due to AL, comparing the first iteration with the last one for every classifier and their fusion, so that we can see that AL is effective since performance has increased for every classifier at the end of the process. Secondly, considering the fusion result, it shows that the improvement is not only due to the increased size of training set but also to the chosen sample selection strategy. The image underlines in-

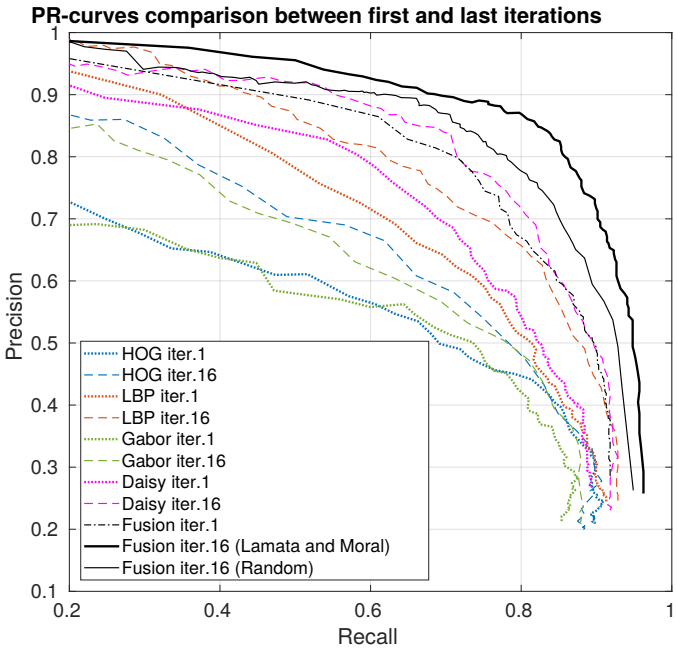


Figure 9: PR curves for the individual classifiers, as well as the fusion between them, for the first and the last iterations. For the sample selection, we compare Lamata and Moral's strategy with the random selector (which benefits only from a larger training set).

deed the consistent gap between the two fusion results at the last iteration, which corresponds to random sample selector and maximum entropy sample selector (considering Lamata and Moral's entropy definition). This fact underlines the importance of having defined an adapted fusion strategy which is able to take into account imprecision while at the same time providing clues for the AL process.

5. Conclusion

Our work proposes a belief function based fusion strategy for the application of pedestrian detection in high-density crowds. The algorithm relies on a pool of heterogeneous detectors, and it exhibits two fundamental advantages with respect to the many existing ensemble methods. Firstly, given the reduced size of the targets, we take into account jointly the imprecision related to the individual decisions during the calibration process along with the detection imprecision exhibited in the image space. Secondly, we exploit the disagreement among the individual detectors in order to add new samples to the training set in an optimal manner with respect to the information gain evaluated with different evidential entropy measures. The results highlight the effectiveness of the detector fusion and confirm the expected behaviors of the different entropy measures we considered in our study.

Even if in our application we considered only two classes, the method could be easily extended to handle more than two singleton hypotheses: the proposed Maximum Entropy and Maximum Conflict criteria would naturally apply, while the Maximum Ignorance criterion will possibly benefit from distinction between partial ignorance values. Refining conflict analysis, we could also take into account the difference between external and internal conflict components, to find a balance between exploration and exploitation of the version space.

From a methodological point of view, the perspectives of this work are related to extending it for performing neural network late fusion, for which the widely used combinations do not rely on an evidential estimation of the imprecision or disagreement (i.e. [76, 77, 78]). Application wise, the proposed strategy is well adapted in contexts in which it is desirable to pinpoint efficiently new training samples from a large pool

of heterogeneous data based on the lack of consensus among many classifiers. For the specific field of crowd analysis, we are also interested in performing local density estimation which could be helpful in identifying instability areas without the need to perform accurate individual detections.

References

[1] P. Xu, F. Davoine, T. Denoeux, Evidential combination of pedestrian detectors, in: British Machine Vision Conference, 2014, pp. 1–14.

[2] A. Ziebinski, R. Cupek, H. Erdogan, S. Waechter, A survey of ADAS technologies for the future perspective of sensor fusion, in: International Conference on Computational Collective Intelligence, Springer, 2016, pp. 135–146.

[3] Z. Zhang, C. Conly, V. Athitsos, A survey on vision-based fall detection, in: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, ACM, 2015, pp. 46:1–46:7.

[4] X. Wang, M. Wang, W. Li, Scene-specific pedestrian detection for static video surveillance, IEEE transactions on pattern analysis and machine intelligence 36 (2) (2014) 361–374.

[5] P. Minary, F. Pichon, D. Mercier, E. Lefevre, B. Droit, Face pixel detection using evidential calibration and fusion, International Journal of Approximate Reasoning 91 (2017) 202–215.

[6] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, S. Yan, Crowded scene analysis: A survey, IEEE Trans. Circuits Syst. Video Techn. 25 (3) (2015) 367–386.

[7] R. Benenson, M. Omran, J. Hosang, B. Schiele, Ten years of pedestrian detection, what have we learned?, in: European Conference on Computer Vision, Springer, 2014, pp. 613–627.

[8] S. Zhang, R. Benenson, M. Omran, J. Hosang, B. Schiele, Towards reaching human performance in pedestrian detection, IEEE transactions on pattern analysis and machine intelligence 40 (4) (2018) 973–986.

[9] X. Wang, Intelligent multi-camera video surveillance: A review, Pattern recognition letters 34 (1) (2013) 3–19.

[10] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR, IEEE, 2005, pp. 886–893.

[11] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern recognition 29 (1) (1996) 51–59.

[12] D. Cohn, L. Atlas, R. Ladner, Improving generalization with active learning, Machine learning 15 (2) (1994) 201–221.

[13] B. Settles, Active learning, Synthesis Lectures on Artificial Intelligence and Machine Learning 6 (1) (2012) 1–114.

[14] D. D. Lewis, W. A. Gale, A sequential algorithm for training text classifiers, in: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Springer-Verlag New York, Inc., 1994, pp. 3–12.

[15] Z. Xu, K. Yu, V. Tresp, X. Xu, J. Wang, Representative sampling for text classification using support vector machines, in: European Conference on Information Retrieval, Springer, 2003, pp. 393–407.

[16] C. Chao, M. Cakmak, A. L. Thomaz, Transparent active learning for robots, in: Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, IEEE, 2010, pp. 317–324.

[17] S. C. Hoi, R. Jin, J. Zhu, M. R. Lyu, Batch mode active learning and its application to medical image classification, in: Proceedings of the 23rd international conference on Machine learning, ACM, 2006, pp. 417–424.

[18] N. Cebron, M. R. Berthold, Active learning for object classification: from exploration to exploitation, Data Mining and Knowledge Discovery 18 (2) (2009) 283–299.

[19] D. D. Lewis, J. Catlett, Heterogeneous uncertainty sampling for supervised learning, in: Proceedings of the eleventh international conference on machine learning, 1994, pp. 148–156.

[20] G. Schohn, D. Cohn, Less is more: Active learning with support vector machines, in: ICML, Citeseer, 2000, pp. 839–846.

[21] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, Journal of machine learning research 2 (Nov) (2001) 45–66.

[22] P. Donmez, J. G. Carbonell, P. N. Bennett, Dual strategy active learning, in: European Conference on Machine Learning, Springer, 2007, pp. 116–127.

[23] S.-J. Huang, R. Jin, Z.-H. Zhou, Active learning by querying informative and representative examples, in: Advances in neural information processing systems, 2010, pp. 892–900.

[24] K. Brinker, Incorporating diversity in active learning with support vector machines, in: ICML, 2003, pp. 59–66.

[25] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, W. J. Emery, SVM active learning approach for image classification using spatial information, IEEE Transactions on Geoscience and Remote Sensing 52 (4) (2014) 2217–2233.

[26] H. S. Seung, M. Opper, H. Sompolinsky, Query by committee, in: Proceedings of the fifth annual workshop on Computational learning theory, ACM, 1992, pp. 287–294.

[27] L. Breiman, Bagging predictors, Machine learning 24 (2) (1996) 123–140.

[28] Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences 55 (1) (1997) 119–139.

[29] A. K. McCallumzy, K. Nigamy, Employing EM and pool-based active learning for text classification, in: Proc. International Conference on Machine Learning (ICML), Citeseer, 1998, pp. 359–367.

[30] S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1) (1951) 79–86.

[31] P. Melville, S. M. Yang, M. Saar-Tsechansky, R. Mooney, Active learning for probability estimation using Jensen-Shannon divergence, in: European Conference on Machine Learning, Springer, 2005, pp. 268–279.

[32] G. Ngai, D. Yarowsky, Rule writing or annotation: Cost-efficient resource usage for base noun phrase chunking, in: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2000, pp. 117–125.

[33] Y. Zhao, C. Xu, Y. Cao, Research on query-by-committee method of active learning and application, in: International Conference on Advanced Data Mining and Applications, Springer, 2006, pp. 985–991.

[34] S. Kee, E. del Castillo, G. Runger, Query-by-committee improvement with diversity and density in batch active learning, Information Sciences 454 (2018) 401–418.

[35] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, Journal of Machine Learning Research 12 (Jul) (2011) 2211–2268.

[36] L. Ma, S. Destercke, Y. Wang, Online active learning of decision trees with evidential data, Pattern Recognition 52 (2016) 33–45.

[37] T. Gao, D. Koller, Active classification based on value of classifier, in: Advances in Neural Information Processing Systems, 2011, pp. 1062–1070.

[38] T. Reineking, Active classification using belief functions and information gain maximization, International Journal of Approximate Reasoning 72 (2016) 43–54.

[39] J. Vandoni, S. Le Hégarat-Mascle, E. Aldea, Belief function definition for ensemble methods - application to pedestrian detection in dense crowds, in: Proceedings of the 21st International Conference on Information Fusion (FUSION), 2018, pp. 2481–2488.

[40] G. Shafer, A mathematical theory of evidence, Vol. 1, Princeton university press Princeton, 1976.

[41] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (2) (1994) 191–234.

[42] T. Denoeux, 40 years of Dempster-Shafer theory, International Journal of Approximate Reasoning 79 (2016) 1–6.

[43] A. Kallel, S. L. Hégarat-Mascle, Combination of partially non-distinct beliefs: The cautious-adaptive rule, Int. J. Approx. Reasoning 50 (7) (2009) 1000–1021.

[44] A. Jousselme, P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, Int. J. Approx. Reasoning 53 (2) (2012) 118–145.

[45] M. Lachaize, S. Le Hégarat-Mascle, E. Aldea, A. Maitrot, R. Reynaud, Evidential framework for error correcting output code classification, Engineering Applications of Artificial Intelligence 73 (2018) 10–21.

[46] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Advances in large margin classifiers 10 (3) (1999) 61–74.

[47] P. Xu, F. Davoine, H. Zha, T. Denoeux, Evidential calibration of binary SVM classifiers, International Journal of Approximate Reasoning 72 (2016) 55–70.

[48] I. Bloch, Defining belief functions using mathematical morphology–application to image fusion under imprecision, Int. journal of approximate reasoning 48 (2) (2008) 437–465.

[49] J. Vandoni, E. Aldea, S. Le Hégarat-Mascle, An evidential framework for pedestrian detection in high-density crowds, in: Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on, IEEE, 2017, pp. 1–6.

[50] J. Vandoni, E. Aldea, S. Le Hégarat-Mascle, Active learning for high-density crowd count regression, in: Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on, IEEE, 2017, pp. 1–6.

[51] U. Höhle, Entropy with respect to plausibility measures, in: Proceedings of the 12th IEEE international symposium on multiple-valued logic, 1982, pp. 167–169.

[52] R. R. Yager, Entropy and specificity in a mathematical theory of evidence, International Journal of General System 9 (4) (1983) 249–260.

[53] H. T. Nguyen, On entropy of random sets and possibility distributions, The Analysis of Fuzzy Information 1 (1987) 145–156.

[54] N. R. Pal, J. C. Bezdek, R. Hemasinha, Uncertainty measures for evidential reasoning I: A review, International Journal of Approximate Reasoning 7 (3-4) (1992) 165–183.

[55] N. R. Pal, J. C. Bezdek, R. Hemasinha, Uncertainty measures for evidential reasoning II: A new measure of total uncertainty, International Journal of Approximate Reasoning 8 (1) (1993) 1–16.

[56] D. Dubois, H. Prade, Properties of measures of information in evidence and possibility theories, Fuzzy sets and systems 24 (2) (1987) 161–182.

[57] M. T. Lamata, S. Moral, Measures of entropy in the theory of evidence, International Journal Of General System 14 (4) (1988) 297–305.

[58] Y. Deng, Deng entropy, Chaos, Solitons & Fractals 91 (2016) 549–553.

[59] R. Jiroušek, P. P. Shenoy, A new definition of entropy of belief functions in the Dempster–Shafer theory, International Journal of Approximate Reasoning 92 (2018) 49–65.

[60] A.-L. Jousselme, C. Liu, D. Grenier, É. Bossé, Measuring ambiguity in the evidence theory, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 36 (5) (2006) 890–903.

[61] B. R. Cobb, P. P. Shenoy, On the plausibility transformation method for translating belief function models to probability models, International journal of approximate reasoning 41 (3) (2006) 314–330.

[62] J. Dezert, F. Smarandache, A. Tchamova, On the Blackman's association problem, in: Proceedings of the 6th Annual Conference on Information Fusion, 2003, pp. 1371–1378.

[63] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, Information and Security 9 (2002) 13–57.

[64] J. Dezert, F. Smarandache, Presentation of DSmT, in: Advances and Applications of DSmT for Information Fusion, American Research Press, 2004, pp. 3–35.

[65] M. Daniel, Conflicts within and between belief functions, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2010, pp. 696–705.

[66] M. Daniel, Non-conflicting and conflicting parts of belief functions, in: ISIPTA, Vol. 11, Citeseer, 2011, pp. 149–158.

[67] S. Destercke, T. Burger, Toward an axiomatic definition of conflict between belief functions, IEEE transactions on cybernetics 43 (2) (2013) 585–596.

[68] E. Lughofer, Single-pass active learning with conflict and ignorance, Evolving Systems 3 (4) (2012) 251–271.

[69] M. Sharma, M. Bilgic, Evidence-based uncertainty sampling for active learning, Data Mining and Knowledge Discovery 31 (1) (2017) 164–202.

[70] E. Aldea, K. H. Kiyani, Hybrid focal stereo networks for pattern analysis in homogeneous scenes, in: Computer Vision - ACCV 2014 Workshops - Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part III, 2014, pp. 695–710.

[71] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, 2001, pp. 511–518.

[72] M. Li, S. Bao, W. Dong, Y. Wang, Z. Su, Head-shoulder based gender recognition, in: ICIP, 2013, pp. 2753–2756.

[73] P. Li, G. Samorodnitsk, J. Hopcroft, Sign Cauchy projections and Chi-Square kernel, in: Advances in Neural Information Processing Systems, 2013, pp. 2571–2579.

[74] E. Tola, V. Lepetit, P. Fua, Daisy: An efficient dense descriptor applied to wide-baseline stereo, IEEE TPAMI 32 (5) (2010) 815–830.

[75] M. E. Ramirez-Loaiza, M. Sharma, G. Kumar, M. Bilgic, Active learning: an empirical study of common baselines, Data Mining and Knowledge Discovery 31 (2) (2017) 287–313.

[76] C. Feichtenhofer, A. Pinz, A. Zisserman, Convolutional two-stream network fusion for video action recognition, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1933–1941.

[77] H. Hu, Z. Wang, J.-Y. Lee, Z. Lin, G.-J. Qi, Temporal domain neural encoder for video representation learning, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, IEEE, 2017, pp. 2192–2199.

[78] Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, S.-F. Chang, Exploiting feature and class relationships in video categorization with regularized deep neural networks, IEEE transactions on pattern analysis and machine intelligence 40 (2) (2018) 352–364.