

# Bi-capacity Choquet Integral for Sensor Fusion with Label Uncertainty

Hersh Vakharia

*University of Michigan*

Ann Arbor, MI

hershv@umich.edu

Xiaoxiao Du

*University of Michigan*

Ann Arbor, MI

xiaodu@umich.edu

**Abstract**—Sensor fusion combines data from multiple sensor sources to improve reliability, robustness, and accuracy of data interpretation. The Fuzzy Integral (FI), in particular, the Choquet integral (ChI), is often used as a powerful nonlinear aggregator for fusion across multiple sensors. However, existing supervised ChI learning algorithms typically require precise training labels for each input data point, which can be difficult or impossible to obtain. Additionally, prior work on ChI fusion is often based only on the normalized fuzzy measures, which bounds the fuzzy measure values between  $[0, 1]$ . This can be limiting in cases where the underlying scales of input data sources are bipolar (i.e., between  $[-1, 1]$ ). To address these challenges, this paper proposes a novel Choquet integral-based fusion framework, named *Bi-MIChI* (pronounced “bi-mi-kee”), which uses bi-capacities to represent the interactions between pairs of subsets of the input sensor sources on a bi-polar scale. This allows for extended non-linear interactions between the sensor sources and can lead to interesting fusion results. Bi-MIChI also addresses label uncertainty through Multiple Instance Learning, where training labels are applied to “bags” (sets) of data instead of per-instance. Our proposed Bi-MIChI framework shows effective classification and detection performance on both synthetic and real-world experiments for sensor fusion with label uncertainty. We also provide detailed analyses on the behavior of the fuzzy measures to demonstrate our fusion process.

**Index Terms**—bi-capacity, choquet integral, fuzzy measures, sensor fusion, label uncertainty, classification

## I. INTRODUCTION

Sensors are all around us collecting data. Each sensor may provide complementary or reinforcing information that supports tasks such as target detection, classification, or scene understanding. The fuzzy integral (FI), in particular, the Choquet integral (ChI), has been used widely as a powerful nonlinear aggregator for sensor fusion [1]–[5]. The ChI is based on fuzzy measures, or “capacities”, which are a set of real-valued parameters that represent the interactions between input sources and help determine the decision-making process.

Suppose we are fusing  $m$  sensor data sources,  $C = \{c_1, c_2, \dots, c_m\}$ . The power set of  $C$  is denoted as  $2^C$ , which contains all possible crisp subsets of  $C$ . The Choquet integral fusion relies on a set of fuzzy measures (length  $2^C$ ), where each fuzzy measure element, or “capacity”, correspond to each subset of the sensor source combinations. As an example, if the

This material is based upon work supported by the National Science Foundation under Grant IIS-2153171-CRII: III: Explainable Multi-Source Data Integration with Uncertainty.

fuzzy measure is denoted as  $g$ , the notation  $g_{\{1,2\}}$  indicates the contributions from the intersection of sensor input source 1 and source 2.

Two challenges exist for Choquet integral-based fusion methods in the literature. First, most prior work on ChI fusion adopt the normalized fuzzy measures [6]–[10], which bounds the fuzzy measures between 0 and 1. To put it mathematically, the normalized fuzzy measure,  $g$ , is a real-valued function that maps  $2^C \rightarrow [0, 1]$ . This means, any combinations of the input sources are always weighted non-negatively. This can be troublesome as the normalized fuzzy measure may overlook background information provided by the sensor sources. Here is an example to illustrate this effect. Consider a binary classification application where the ChI fusion is performed to detect humans/pedestrians in a scene. We want the human/pedestrian pixels to have target label “1” and everything else (background trees, buildings, and shadows, etc.) to have non-target label “0”. Assume one of the input sensors is a shadow sensor that can detect shadows well (therefore, this sensor source has high values on shadow region and low values on “human” pixels). If the normalized fuzzy measure is used, it will assign value 0 to the measure element associated with the shadow sensor (in other words, ignore the shadow sensor input completely), as it did not contribute to the positive target (human) class. However, we argue that the shadow sensor also carries useful information that should not be ignored. If we know where the shadows are, we can apply a negative weight on the shadows to mark the “non-human” regions, and potentially improve the accuracy and remove shadow artifact from the pedestrian detection results. Thus, it is important to explore extended fuzzy measures to accommodate and incorporate useful and complementary information from all sensor sources. In this work, we propose the use of bi-capacities [11], [12], which are a generalization of fuzzy measures/capacities that are useful when data follows a bi-polar scale of  $[-1, 1]$ . The bi-capacities naturally extend and generalize the ChI fusion to produce fusion results on a bipolar scale and can result in improved detection performance (as shown later in Experiments).

Second, existing supervised ChI fusion algorithms often require precise instance-level labels for training. However, accurate labels can often be difficult, if not impossible, to obtain. Let us consider the pedestrian detection application again. It can be tedious and expensive to label each pixel on

an image, but it is very simple to draw a bounding box region around where a person may appear. Based on the Multiple Instance Learning (MIL) framework [13], we name such a region a “bag”, where each bag contains a set of pixels. The MIL framework has been explored in previous literature for fuzzy fusion with bag-level label uncertainties [8], [14]. However, none of these work handles fuzzy measures outside the  $[0, 1]$  range. This work is the first, to our knowledge, that addresses bi-capacity ChI fusion with bipolar sensor inputs and label uncertainty under the MIL framework.

The remainder of the paper is organized as follows. Section II provides a review on related works and discusses prior work in ChI fusion and MIL formulation. Section III presents the definition and properties of the bi-capacity ChI formulations. Section IV describes the proposed Bi-MICChI algorithm, its objective functions, and optimization strategies. Section V presents experimental results of the proposed algorithm on both simulated and real datasets for multi-sensor fusion. Section VI discusses the main findings, conclusions, and future work.

## II. RELATED WORK

### A. Bi-capacity & Choquet Integral Applications

Bi-capacities and bipolar capacities were developed in decision theory where the underlying scales are bipolar [11]. They extend the concept of capacities, or fuzzy measures, by representing interactions of pairs of subsets on a bipolar scale, which can model multidimensional and complex problems. The bi-capacities and bipolar ChI has been explored previously in [12], [15], [16]. However, these works mainly use hand-crafted, synthetic examples as an illustration and did not handle data and label uncertainties in real sensing data. To our knowledge, no prior experiments using bi-capacities have been conducted on real data integration applications considering imprecise labels.

Bi-capacities have successfully been applied in multi-criteria decision making under uncertainty, where they can be used to aggregate positive and negative preferences [17]–[19]. Bi-capacities have also been used to perform complex, dynamic evaluations, such as evaluating regional eco-efficiency in [20]. However, they have yet to be applied in the realm of sensor fusion as a way to represent the complex relationship between the input sources to be fused.

The Choquet Integral, on the other hand, has been used as an effective nonlinear aggregation operation for many applications, such as landmine detection [2], [21], soil-type classification [22], gesture recognition [23], and pedestrian detection [24]. However, these applications do not utilize bi-capacities for modelling the complementary relationships between sources on a bipolar scale. This work will extend the ChI fusion to a bipolar scale using bi-capacities.

### B. Multiple Instance Learning

The Multiple Instance Learning (MIL) framework was created to address the problem of label ambiguity and uncertainty in supervised learning [13]. Instead of requiring instance-level

labels for supervised learning, MIL allows for uncertain labels for “bags” of instances. Bags are labeled positive if it contains at least one positive (“target”) instance and negative if all instances are negative (“non-target”). It has been successfully applied to a variety of applications, such as medical image and video analysis [25], landmine detection [26]–[28], scene segmentation [29], [30], and other remote sensing tasks [31]. This work aims to extend MIL for bi-capacity ChI sensor fusion to handle label uncertainties, where the input labels are at bag-level, instead of instance-level (per pixel) labels.

### C. Multiple Instance Choquet Integral

This work extends the work of Multiple Instance Choquet Integral (MICI) for classification and regression [7], [8], which is a supervised ChI fusion framework under the MIL framework. One limitation of MICI is that it utilizes normalized fuzzy measures bounded between  $[0, 1]$  and typically requires the input sensor data to be normalized between  $[0, 1]$  as well. While MICI’s normalized fuzzy measures can model some interactions between sources, the use of bi-capacities in this paper introduces new functionalities where some sensor source combinations can now be weighed negatively against others (which was not done before). This also leads to improved classification performance and enhanced interpretability. We also developed a new objective function to train the proposed Bi-MICChI algorithm and learn the bi-capacity values.

## III. BI-CAPACITY CHI DEFINITIONS

Bi-capacities are a variety of capacities, or fuzzy measures, that are useful when data follows a bipolar scale [11]. When applied to fusion, they allow for interesting results where some sources, or combinations of sources, can be weighed negatively against others. Let  $C = \{c_1, c_2, \dots, c_m\}$  denote the  $m$  sensor sources to be fused. A monotonic bi-capacity,  $g$ , is a real valued function that maps  $S(C) \rightarrow [-1, 1]$ , where  $S(C) = \{(A, B) : A \subseteq C, B \subseteq C, A \cap B = \emptyset\}$ .  $S(C)$  represents pairs of disjoint, crisp, subsets of  $C$ . It satisfies the following properties [12]:

1) Boundary Conditions:

$$g(\emptyset, \emptyset) = 0, g(C, \emptyset) = 1, g(\emptyset, C) = -1$$

2) Monotonicity:

$$\text{If } A \subseteq E \text{ and } B \supseteq F, \text{ then } g(A, B) \geq g(E, F)$$

If we compare these properties to those of the normalized fuzzy measure (e.g., see Definition 2.1 of [32]), we note the following specific properties for bi-capacities. First, the bi-capacities capture information of *ternary* importance from multi-source data and have the potential of extracting useful and complementary information on bipolar scales. The sensor set  $C$  can be configured into  $3^m$  disjoint subset pairs, with  $3^m - 3$  non-boundary, learnable elements. Each element of the bi-capacity represents a weighting of the first subset “against” the second subset. This property allows for interesting fusion results, where some sources, or sets of sources, can be weighed negatively against other sources. Additionally, sources that

have negative information can still be used to reinforce the final fusion result. In this paper, bi-capacity elements will be denoted with a subscript of its corresponding pair of subsets. For example,  $g_{1,23}$  corresponds to subset pair  $(\{c_1\}, \{c_2, c_3\})$ . In this example, the value of bi-capacity element  $g_{1,23}$  weighs  $c_1$  against  $c_2$  and  $c_3$ .

Second, to satisfy monotonicity, a bi-capacity *increases* with inclusion in the first subset and *decreases* with inclusion in the second subset. Fig. 1 illustrates the monotonicity property of a bi-capacity for an example with three sources. As shown, bi-capacity elements  $g_{123,0}$  and  $g_{0,123}$  are the uppermost and lowermost bounds of  $+1$  and  $-1$ , respectively. There are various paths between the bounds that satisfy monotonicity; one possible path is highlighted in red. As an example, take bi-capacity elements of  $g_{12,3}$  and  $g_{2,3}$ , which correspond to subset pairs  $(\{c_1, c_2\}, \{c_3\})$  and  $(\{c_2\}, \{c_3\})$ , respectively. The monotonicity property shows that  $g_{12,3}$  is an upper bound of  $g_{2,3}$ , as  $\{c_1, c_2\} \subseteq \{c_2\}$  and  $\{c_3\} \supseteq \{c_3\}$ .

Bi-capacities with three sources can also be represented using a three-dimensional lattice structure, as shown in Fig. 1 in [11]. In this paper, the values of the bi-capacities are represented in matrix form, as shown in tables II and III in the experiments section. As seen in the tables, the bi-capacities create a repeating fractal pattern within the matrix [11].

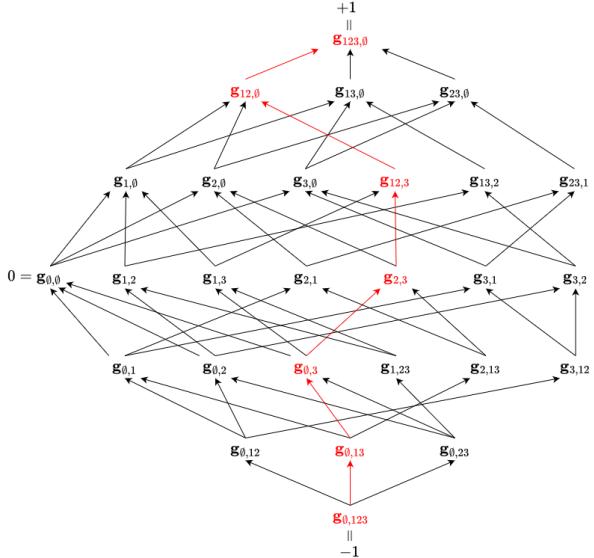


Fig. 1: Illustration of bi-capacity element relationships given three sources. Red shows an example path of monotonicity.

Given a bi-capacity,  $g$ , instance-level fusion is computed with a using the discrete Choquet Integral, which is well-defined for bi-capacities [12]. The bi-capacity Choquet Integral,  $\mathcal{C}_g$ , on instance  $\mathbf{x}_n$  can be computed as

$$\mathcal{C}_g(\mathbf{x}_n) = \sum_{k=1}^m \left[ (|h(c_k; \mathbf{x}_n)| - |h(c_{k-1}; \mathbf{x}_n)|) \cdot g(A_k \cap C^+, A_k \cap C^-) \right], \quad (1)$$

where  $h(c_k; \mathbf{x}_n)$  denotes the output of the  $k^{th}$  sensor source input,  $c_k$ , on the  $n^{th}$  data instance,  $\mathbf{x}_n$ .  $C$  is sorted such that  $|h(c_1; \mathbf{x}_n)| \leq |h(c_2; \mathbf{x}_n)| \leq \dots \leq |h(c_m; \mathbf{x}_n)|$ . Since there are  $m$  classifiers,  $h(c_0; \mathbf{x}_n)$  is defined to be 0. When using the bi-capacity  $g$  with the Choquet Integral, it is defined that  $A_k = \{c_k, \dots, c_m\}$ , where  $C$  is sorted. Additionally,  $C^+ = \{i \in C \mid h(c_i; \mathbf{x}_n) \geq 0\}$  and  $C^- = C \setminus C^+$ .

#### IV. THE PROPOSED BI-MICHI

To utilize the Choquet Integral successfully for fusion, the non-boundary elements of the bi-capacity must be learned from a set of training data. In the MIL framework, data is organized in “bags” with uncertain and imprecise labels. Based on the MIL assumptions, bags are labeled positive if at least one instance in the bag is positive and negative if all instances in that bag are negative [13].

Including or excluding the  $g_{0,0} = 0$  bound can provide different results for the bi-capacity Chi. If this bound is removed, elements of  $g_{0,\cdot}$  can be positive and  $g_{\cdot,0}$  can be negative, which allows for an inversion effect on the data. In the following, we present two variations on the objective functions, one without and one with the  $g_{0,0}$  term, and describe our optimization strategies to learn the bi-capacities from the training data and bag-level labels.

##### A. Objective Function 1: without $g_{0,0} = 0$

The objective function for the negative and positive bags can be written as

$$J^- = \sum_{a=1}^{B^-} \max_{\forall \mathbf{x}_{ai}^- \in \mathcal{B}_a^-} (\mathcal{C}_g(\mathbf{x}_{ai}^-) + 1)^2, \quad (2)$$

$$J^+ = \sum_{b=1}^{B^+} \min_{\forall \mathbf{x}_{bj}^+ \in \mathcal{B}_b^+} (\mathcal{C}_g(\mathbf{x}_{bj}^+) - 1)^2, \quad (3)$$

where  $B^+$  is the number of positive bags,  $B^-$  is the number of negative bags,  $\mathbf{x}_{ai}^-$  is the  $i^{th}$  instance in the  $a^{th}$  negative bag,  $\mathcal{B}_a^-$ , and  $\mathbf{x}_{bj}^+$  is the  $j^{th}$  instance in the  $b^{th}$  positive bag,  $\mathcal{B}_b^+$ .  $\mathcal{C}_g$  is the Choquet Integral output for a bi-capacity,  $g$ .

This objective function is parallel to the min-max model for normalized fuzzy measures [8]. However, the bi-capacities adds additional ternary information (as defined in Section III) and the computation of bi-capacity Chi is different compared with prior work and now has the range between  $[-1, 1]$ . The bi-capacity,  $g$ , is obtained by minimizing the objective function in equation 4. The  $J^-$  term in Eq.(2) encourages the Choquet integral of instances in the negative bags to be “ $-1$ ”, while the  $J^+$  term in Eq.(3) encourages the instances in the positive bags to “ $+1$ ”. Therefore, given training data with positively and negatively labeled bags, minimizing  $J$  results in a bi-capacity that will produce a high confidence on the target and low confidence on the background. The overall objective can be written as

$$\min_g J = J^- + J^+. \quad (4)$$

### B. Objective Function 2: with $\mathbf{g}_{\emptyset, \emptyset} = 0$

Since the  $\mathbf{g}_{\emptyset, \emptyset} = 0$  bound enforces  $\mathbf{g}_{\cdot, \emptyset} \geq 0$  and  $\mathbf{g}_{\emptyset, \cdot} \leq 0$ , the Chi is unable to produce an inversion effect on negative data. Therefore, a different objective function must be used to properly represent the negative data in the fused result. We can define a new set of  $J^+$  and  $J^-$  as

$$J^- = \sum_{a=1}^{B^-} \max_{\forall \mathbf{x}_{ai}^- \in \mathcal{B}_a^-} (\mathcal{C}_{\mathbf{g}}(\mathbf{x}_{ai}^-) - 0)^2, \quad (5)$$

$$J^+ = \sum_{b=1}^{B^+} \min_{\forall \mathbf{x}_{bj}^+ \in \mathcal{B}_b^+} (1 - |\mathcal{C}_{\mathbf{g}}(\mathbf{x}_{bj}^+)|)^2. \quad (6)$$

Instead of having  $J^-$  encourage negative bags to have a Chi output of  $-1$ , we instead encourage them to the neutral value of  $0$ . Additionally,  $J^+$  now encourages data in the positive bags to be close to the poles of  $-1$  and  $+1$ . The idea is that the sources to be fused may detect the target class negatively or positively. Thus, instead of trying to invert the negative detection, we can represent the target as values that are on either pole. Then, after fusion, the absolute value of the result can be taken to reflect high confidence in the target. The overall objective is the sum of these two terms, as defined in Eq. (4).

### C. Optimization Algorithm

The optimization is achieved through the use of an evolutionary algorithm following that of MICI [8], with several modifications based on bi-capacities. Algorithm 1 shows the pseudo-code of the optimization and fusion steps for the proposed Bi-MICHI algorithm, and Table I explains the parameters used. Fig. 2 shows a flowchart of the algorithm. The algorithm requires training data (sensor inputs from multiple sources) and bag-level labels for the training bags as inputs. The output of the algorithm is the learned bi-capacities, which will be used to generate instance-level fusion results given multi-sensor sources.

First, a population (size  $P$ ) of bi-capacities is initialized. The experiments in this paper used  $P = 36$ . Bi-capacities are initialized starting at the  $\mathbf{g}_{\emptyset, \emptyset} = 0$  boundary and moving outward toward the upper and lower bounds. This process is completed similarly to breadth-first-search exploration [33], where the neighbor nodes are the upper and lower bounds of the current node. As an example, when  $\mathbf{g}_{\emptyset, \emptyset}$  is included in Objective Function 2, after starting at  $\mathbf{g}_{\emptyset, \emptyset}$ , the lower bound elements  $\mathbf{g}_{1,0}$ ,  $\mathbf{g}_{2,0}$ , and  $\mathbf{g}_{3,0}$  and upper bound elements  $\mathbf{g}_{0,1}$ ,  $\mathbf{g}_{0,2}$ , and  $\mathbf{g}_{0,3}$  are the “neighbors” that are to be explored and sampled next. The current element is sampled uniformly randomly between the minimum of all the previously sampled upper bounds and the maximum of all the previously sampled lower bounds. This same methodology is used for if the  $\mathbf{g}_{\emptyset, \emptyset} = 0$  boundary is not used (Objective Function 1); the  $0$  bound is simply not enforced during sampling.

In the main optimization loop, bi-capacities are updated using a combination of small-scale and large-scale mutations.

In the small-scale mutation, only one of the bi-capacity elements is re-sampled. The element to be sampled is determined through a multinomial distribution based on the number of times each element of the bi-capacity is used by the Choquet Integral during training. Therefore, the element that is used the most during training will have the highest probability to be resampled. Eq. (7) shows the formula for the probability for resampling, where  $v_{A,B}$  is the number of times bi-capacity element  $\mathbf{g}_{A,B}$  is used during training. The selected element is resampled uniformly between the associated upper and lower bounds. In the large-scale mutation, an entirely new bi-capacity is sampled. The probability of small and large mutations is determined by a user-defined parameter  $\eta$ . Small mutations occur with a probability of  $\eta$ , and large mutations occur with a probability of  $1 - \eta$ . The value  $\eta = 0.8$  was found to be effective empirically during experimentation.

$$P(\mathbf{g}_{A,B}) = \frac{v_{A,B}}{\sum_{\mathbf{g}_{E,F} \in \mathcal{G}} v_{E,F}} \quad (7)$$

After mutation, the fitness of each bi-capacity is computed using the objective function (4). Bi-capacities with improved fitness are saved. This process of mutation and fitness calculation continues for a maximum of  $I$  iterations ( $I = 5000$  in this paper’s experiments) or until certain stopping criteria is met. In our experiments, the stopping criteria is set to be if the new best fitness is within a threshold,  $\mathbf{J}_T$ , of the previous best fitness (in other words, if the fitness does not change any more than the  $\mathbf{J}_T$  threshold). We chose a small threshold value  $\mathbf{J}_T = 0.001$  as it yielded effective convergence results in our experiments.

TABLE I: Algorithm 1 list of parameters.

Parameter	Explanation
$\mathbf{G}$	Population of bi-capacities
$P$	Bi-capacity population size
$\mathbf{J}^*$	Best fitness value
$\mathbf{g}^*$	Best bi-capacity
$\mathbf{J}_P^i$	Fitnesses of population of bi-capacities at iteration $i$
$\mathbf{J}_T$	Fitness threshold for stopping condition $i$
$I$	Maximum optimization iterations
$\eta$	Small-scale mutation probability

## V. EXPERIMENTS

The proposed Bi-MICHI algorithm is applied to a synthetic and a real-world classification task. Experimental results are presented to illustrate the fusion performance of the proposed algorithm.

### A. Synthetic Dataset

The first experiment showcases the fusion abilities on a synthetic dataset. This dataset was constructed to highlight the proposed framework’s ability to utilize negative information using bi-capacities, as well as its ability to handle bag-level imprecise labels. Two sets of tests are run on this dataset, the first without the  $\mathbf{g}_{\emptyset, \emptyset} = 0$  bound using objective functions (2) and (3) (Section IV-A), and the second uses the  $\mathbf{g}_{\emptyset, \emptyset} = 0$  bound and objective functions (5) and (6) (Section IV-B).

## TRAINING

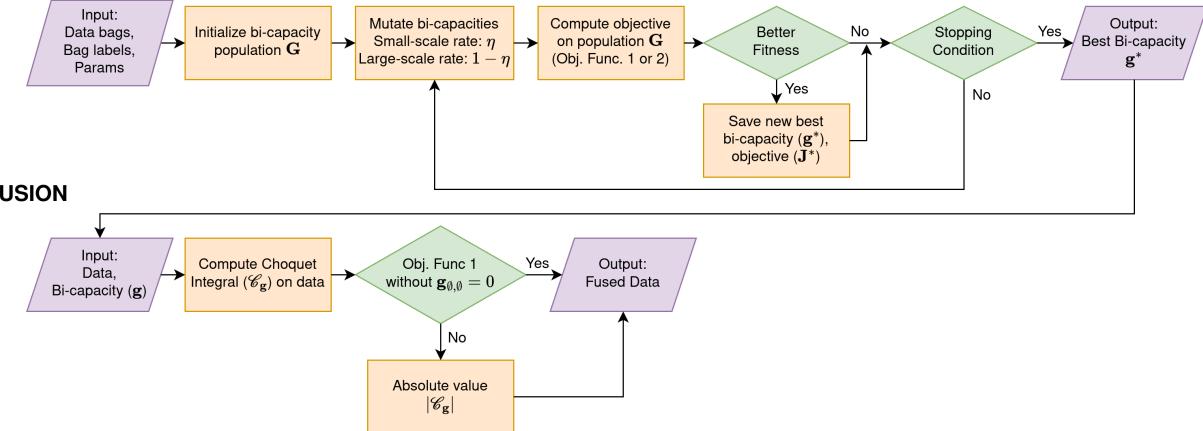


Fig. 2: Flowchart for the Bi-MICHI algorithm.

To generate this simulated example, suppose the data contains a letter “U” and a letter “M” in the scene. We chose these two letter shapes as they are symmetric and easy to visualize. The ground truth is shown in Fig. 4e. Assume we are fusing three sensor inputs to detect both shapes. The first sensor source only detects the letter shape “U” (as shown in Fig. 4a). The second sensor source highlights all background except the letter shape “M” (Fig. 4b). The third sensor source detects all background (non-letter) pixels (Fig. 4c). This example is inspired by real-world target detection applications, where the “U” and “M”-shaped targets consist of different materials, say, the “U”-shape is metal and “M”-shape is plastic, and the goal is to detect both targets. The source 1 detector may only be able to detect the metal target shape “U”, whereas sensor 2 and 3 can only detect vegetation and background (non-target) pixels in the scene.

To generate bag-level data, the SLIC superpixel algorithm [34] is used to group the pixels in the image into bags (superpixels). The generated bags are shown in Fig. 4d. Bags that contain any part of the “U” or “M” targets are labeled positive (target class, highlighted as green), and the remaining bags are labeled as negative (non-target class, shown as red).

1) *Simulated Experiment 1 with Bipolar Labels*: The first test case is shown in Fig. 4. The sources for fusion are shown in Fig. 4a-4c. The goal is to learn a bi-capacity that is able to detect both “U” and “M” shapes with high confidence. In this experiment, bipolar-scale labels are used, where the target class (“U” and “M”) are pushed to label +1 and the non-target class (background) is pushed to label -1, as shown in the ground truth figure (Fig. 4e).

Fig. 4f shows the proposed Bi-MICHI fusion result using Objective Function 1 (bipolar case). We observed some interesting behaviors for the bi-capacities. As shown in Fig. 4f, the “U” shape was detected perfectly (has a positive fusion label of +1), but the “M” shape in the final fusion result was forced to the value -1. This is because the “M” shape in all three sources are not detected, i.e., the input sensor source values

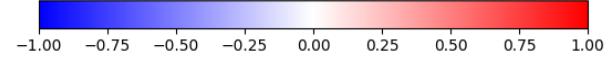


Fig. 3: Colorbar for experimental results.

take the form [-1, -1, -1]. In this case, the Choquet integral pushes all negatives [-1, -1, -1] to the -1 boundary, which resulted in a missed detection for the “M” letter.

Table II shows the learned bi-capacity values. The bi-capacity is notated using the form  $g_{A,B}$ , which represents a weighting of the sensor source subset A “against” the second subset B (see the bi-capacity definition in Section III). The bold elements mark the measure element values that were actually used. Note that not all measure elements were updated based on how the sensor input data is sorted in the ChI computation (see Eq. 1). We observed that, in this experiment, the bi-capacity element  $g_{23,1} = -0.85$ , which represents the combination of sources 2 & 3 both *negatively* detect some parts of the target, while source 1 positively detects the target. The bi-capacity measure element  $g_{23,1}$  has a high magnitude, which shows that the negative contribution of the intersection between sources 2 & 3 *against* the positive contribution of source 1 are weighted highly to help detect the target shapes. This makes sense, as the combination of source 1, the negative of source 2, and the negative of source 3 all contribute to detect both “U” and “M” shapes.

2) *Simulated Experiment 2 with [0, 1] Labels*: The second test case is shown in Fig. 4g, 4h, and 4i. We implement the Object Function 2 with  $g_{0,0} = 0$  and ran Bi-MICHI for fusion. The input sources and bags are the same as the previous experiment. In this test case, however, the ground truth non-target bags are now pushed to a neutral value of 0 (instead of -1 in the previous experiment). The target region (the “U” and “M”) are now pushed to either -1 or +1 (in other words, the absolute value of the target class is pushed to 1). The resulting ground truth map is shown in Fig. 4g, where the “U” and “M” shapes are still highlighted. The ChI fusion result is shown in

---

**Algorithm 1** Bi-MICHI Algorithm

---

**TRAINING**
**Require:** Training Data, Training Labels, Parameters

```

1: Initialize a population of bi-capacities:  $\mathbf{G}$ 
2:  $\mathbf{J}^* = \min(\mathbf{J}_P^0)$ ,  $\mathbf{g}^* = \arg \min_{\mathbf{G}} (\mathbf{J}_P^0)$ 
3: for  $i := 1 \rightarrow I$  do
4:   for  $p := 1 \rightarrow P$  do
5:     Randomly generate  $z \in [0, 1]$ .
6:     if  $z < \eta$  then
7:       Small-scale mutation of  $\mathbf{G}_p$ 
8:     else
9:       Large-scale mutation of  $\mathbf{G}_p$ 
10:    end if
11:   end for
12:   if using  $\mathbf{g}_{\emptyset, \emptyset} = 0$  then
13:     Compute  $\mathbf{G}$  fitness using (5) and (6)     $\triangleright$  Object
14:   else
15:     Compute  $\mathbf{G}$  fitness using (2) and (3)     $\triangleright$  Object
16:   end if
17:   Compute  $\mathbf{J}_d = |\min(\mathbf{J}_P^i) - \mathbf{J}^*|$ .
18:   if  $\min(\mathbf{F}_P^i) < \mathbf{J}^*$  then
19:      $\mathbf{J}^* = \max(\mathbf{J}_P^i)$ ,  $\mathbf{g}^* = \arg \min_{\mathbf{G}} (\mathbf{J}_P^i)$ .
20:   end if
21:   if  $\mathbf{J}_d \leq \mathbf{J}_T$  then break;
22:   end if
23: end for
24: return  $\mathbf{g}^*$ 

```

**RUN FUSION**
**Require:** Data, Bi-capacity  $\mathbf{g}^*$ 

```

24: if using  $\mathbf{g}_{\emptyset, \emptyset} = 0$  then     $\triangleright$  Object Function 2
25:   return  $|\mathcal{C}_g(\text{Data})|$ 
26: else                                 $\triangleright$  Object Function 1
27:   return  $\mathcal{C}_g(\text{Data})$ 
28: end if

```

---

TABLE II: The learned bi-capacity with Simulated Experiment 1 (bipolar labels). Bold elements are updated during Chi fusion.

UM Experiment 1 Bi-capacity, $\mathbf{g}_{A,B}$									
$\backslash$	B	$\emptyset$	1	2	12	3	13	23	123
A	$\emptyset$	0	-0.96	-0.94	-1.00	<b>-0.87</b>	-0.98	<b>-1.00</b>	<b>-1.00</b>
1	1	0.45		-0.91		-0.10		-0.95	
2	2	0.55	-0.94		<b>0.39</b>	-0.96			
12	12	1.00			<b>0.97</b>				
3	3	<b>0.10</b>	-0.86	-0.82	-0.88				
13	13	0.73		-0.72					
23	23	<b>0.77</b>	<b>-0.85</b>						
123	123	1.00							

Fig. 4h, and the absolute value result is shown in Fig. 4i. As seen in the figure, the fused result is able to detect the “UM” with high confidence, showing that pushing positive bags to the  $-1$  and  $+1$  poles can be more successful.

Table III shows the learned bi-capacity values for UM

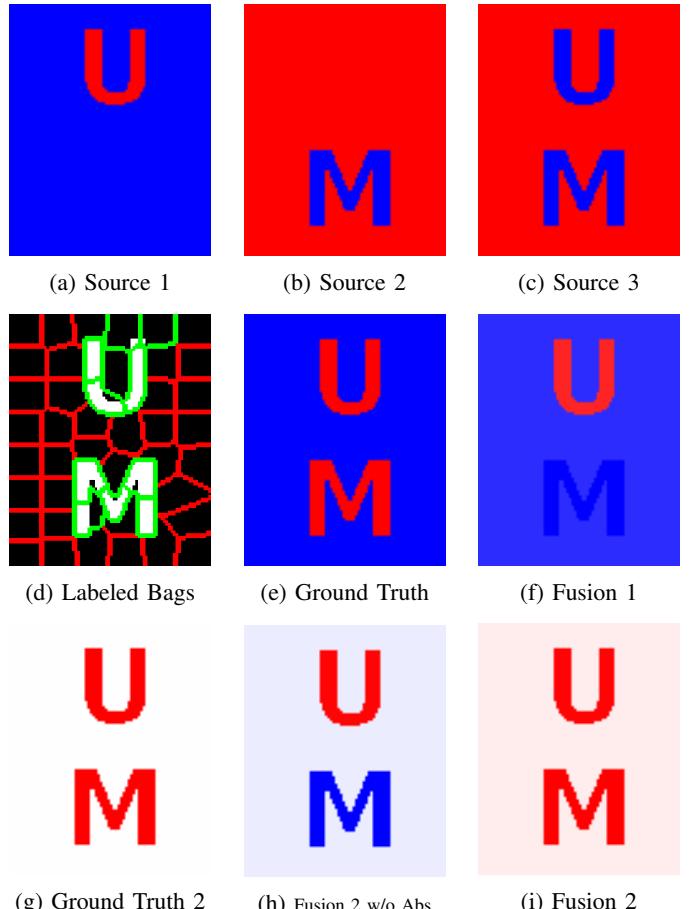


Fig. 4: The Synthetic Experiment results. (a)-(c) are the sources for fusion, (d) shows positive bags in green and negative bags in red, (e) shows the ground truth on the bipolar scale, and (e) shows the fusion result of our Bi-MICHI using objective function 1. (g) shows the ground truth where the non-negative is neutral (label 0). (h) shows the fusion results of our Bi-MICHI using objective function 2, but without the absolute value on the Bi-capacity Choquet term in Eq. (6). (i) shows the fusion results of our Bi-MICHI using the complete objective function 2. All images follow the colorbar in Fig. 3.

experiment 2. In contrast to experiment 1, we observe that the bi-capacity element  $\mathbf{g}_{23,1} = 0.07$ . The magnitude of this element is close to 0, which has the effect of zeroing out the negative contribution of the intersection of sources 2 & 3 against the positive contribution of source 1. This makes sense, as it contributes to the effect of pushing background information to 0. Furthermore, as previously stated, the “M” data of the three sources takes the form  $[-1, -1, -1]$ , which the Chi pushes to  $-1$ . However, taking the absolute value allows for a high confidence detection on the “M” as well.

#### B. Low-Light Pedestrian Detection

We conduct additional experiments on a real-world classification task using the KAIST Multispectral Pedestrian Detection Benchmark [35]. The KAIST dataset contains RGB

TABLE III: The learned bi-capacities with Simulated Experiment 2 ( $[0, 1]$  labels).

		UM Experiment 2 Bi-capacity, $g_{A,B}$							
		$\emptyset$	1	2	12	3	13	23	123
A	$\emptyset$	0	-0.15	-0.09	-0.33	<b>-0.62</b>	-0.89	<b>-0.97</b>	<b>-1.00</b>
	1	0.92		0.81		0.90		0.80	
2	0.83	0.03			<b>0.31</b>	0.02			
12	0.99				<b>0.99</b>				
3	<b>0.91</b>	-0.04	0.84	-0.04					
13	0.95		0.84						
23	<b>0.96</b>	<b>0.07</b>							
123	1.00								

and thermal sensors and the goal is to detect pedestrians in the scene. We selected a low-light setting to illustrate the necessity for sensor fusion, as the thermal camera will be able to highlight pedestrians in low-light whereas the RGB camera will contribute negatively to the detection (RGB cameras will show dark backgrounds with some spots from light sources such as street lamps). Fig. 5 shows an illustration of an image from the KAIST dataset. As shown, under the low light conditions, pedestrians are very difficult to detect, and it is necessary to fuse information from both RGB and thermal cameras. Additionally, since pedestrians are very difficult to see, it is almost impossible to annotate and produce precise pixel-level training labels. However, it is possible to generate bounding boxes (as shown in green in Fig. 5) to indicate possible pedestrian locations. In this experiment, we use the green bounding boxes as the bag-level labels to train our Bi-MICHI algorithm for RGB and thermal fusion.



Fig. 5: KAIST Pedestrian RGB and thermal image pair with ground truth bounding boxes. The images containing the green boxes are over exposed for easier viewing.

The low-light conditions also motivate the use of the bi-capacity's negative weighting. While the thermal image is able to produce high confidence on the pedestrians, there are false positive readings on the background buildings as well as lights. Conversely, the RGB image shows low confidence on the pedestrians, but a high confidence on well-lit areas, like the lights, foreground, and street cones. Negatively weighing the RGB image can contribute to a high confidence on the pedestrians, without relying too heavily on the thermal data.

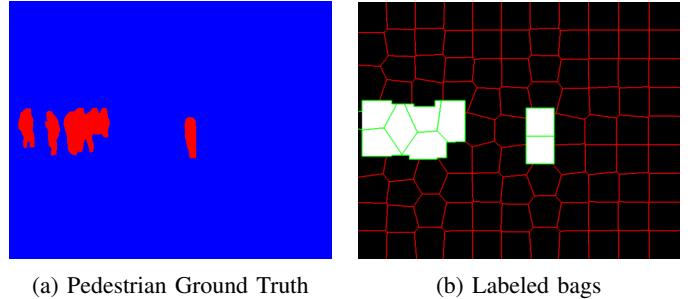


Fig. 6: Pedestrian ground truth and labeled bags. Positive bags are shown in green, and negative bags are shown in red. The bags were generated from the bounding box masks.

The SLIC algorithm [34] is used to generate superpixels (“bags”) for training. Since the dataset provides ground truth in the form of bounding boxes, bags that contain points that intersect with any of the bounding boxes are considered positive, as shown in Fig. 6b. Fig. 6a shows the pixel-level ground truth labels. Note that the pixel-level ground truth labels were not used during training, and were only used to evaluate our final fusion result.

Fig. 7 shows the three sensor sources that are used as fusion inputs, as well as comparison results for fusion. Source 1 (Fig. 7a) utilizes thresholding on a grayscaled version of the RGB image to negatively detect bright areas in the image on the visual spectrum. Source 2 (Fig. 7b) is a grayscaled version of the RGB image with histogram equalization. This shows low, negative confidence on pedestrians and trees, and high confidence on the lights, traffic cones, shadows, and other road elements. Source 3 (Fig. 7c) comes from the thermal data with gamma correction. This source shows high confidence on pedestrians, but there are a few false positives in the background that must be reduced.

The proposed Bi-MICHI was used to perform fusion on this dataset using both Objective Function 1 and Objective Function 2. We compare the fusion results to a variety of comparison methods. First, the min, max, and mean is taken across three sources as a simple fusion baseline. A weighted mean was calculated using weights determined by minimizing the mean squared error between the fusion result and the ground truth. Then, K-nearest neighbors (KNN) with  $k = 100$  and Support Vector Machine (SVM) approaches were used to perform pixel-level fusion. We also compare it to a pre-trained Mask R-CNN [36], which is a neural network-based object detector. Mask R-CNN was run on the low-light RGB image and the segmentation mask is of detected elements of the “person” class. While Mask R-CNN has seen great success in computer vision applications, pre-trained models can exhibit poor performance in low-light applications where object details are unclear without significant image processing. We also ran comparisons with other Choquet Integral-based methods. CI-QP [1] learns a fuzzy measure for fusion by formulating the problem as a quadratic program with instance-level labels. MICI is a previous MIL method for fusion but

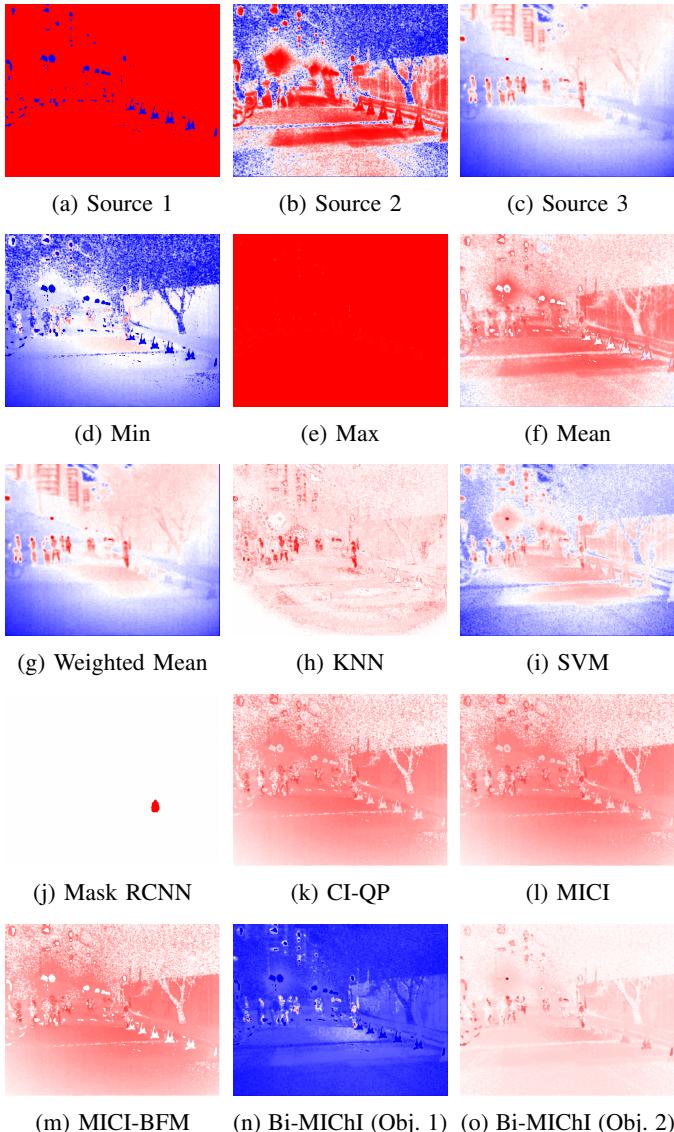


Fig. 7: Pedestrian detection fusion results. All results follow the colorbar in Fig. 3. Best viewed in color.

uses normalized fuzzy measures with elements values in the  $[0, 1]$  range. MICI-BFM [37] is a version of the MICI that uses binary fuzzy measures, which are fuzzy measure with elements that only take values  $\{0, 1\}$ . KNN, SVM, and CI-QP all require instance level labels, while MICI, MICI-BFM, and Bi-MICHI operate on bag-level labels.

Visual results for all methods are shown in Fig. 7. It is important to note that all results are the direct output of the fusion algorithm - some output a confidence map on the  $[-1, 1]$  scale, while others output on a  $[0, 1]$  scale. Naively taking the min, max, and mean of the sources produces poor qualitative results, as the sources highlight conflicting information (background vs. person). The weighted mean produced a result that is very similar to source 3 - this is because the resulting weights heavily favored the third source. KNN highlighted the pedestrians fairly well, but there is a considerable amount of

TABLE IV: Pedestrian detection bi-capacity learned with objective function 1. Bolded elements are used in the Bi-MICHI.

		Pedestrian Detection Bi-capacity (Obj. 1), $g_{A,B}$							
		$\emptyset$	1	2	12	3	13	23	123
A \ B	$\emptyset$	0	<b>-1.00</b>	<b>-1.00</b>	-1.00	<b>-1.00</b>	-1.00	<b>-1.00</b>	-1.00
	1	<b>-0.82</b>		<b>-0.95</b>		<b>-0.92</b>			<b>-0.98</b>
2		<b>-0.99</b>		<b>-1.00</b>			<b>-0.99</b>		<b>-1.00</b>
12		<b>-0.70</b>						<b>-0.79</b>	
3		<b>-1.00</b>		-1.00	<b>-1.00</b>		-1.00		
13		<b>-0.21</b>			<b>-0.59</b>				
23		<b>-0.99</b>		<b>-1.00</b>					
123		<b>1.00</b>							

background noise. The KNN method is also highly sensitive to the parameter choice of  $k$ . When we reduced the  $k$  value (e.g., when  $k = 5$ ), the fusion performance deteriorated significantly and the output became much noisier. Similarly, the SVM approach was able to highlight some pedestrians, but also showed some false positives that occur in the source images. Mask R-CNN falsely identifies a traffic cone as a pedestrian, showing its limitations in low-light conditions. CI-QP, MICI, and MICI-BFM all produce similar results, all of which captured background information with fairly high confidence. Qualitatively, we can see that the proposed Bi-MICHI with objective function 1 is effective at eliminating background information, while still detecting the pedestrians correctly. Bi-MICHI with objective function 2 produces similar results, but contains more false positives compared to objective function 1.

The Bi-MICHI bi-capacities learned with objective function 1 and 2 are shown in Tables IV and V, respectively. We can observe that many of elements in the bi-capacity learned with objective function 1 are set to  $-1$ . This occurs to encourage the background elements to have an output of  $-1$ . The positive detections are the result of the upper boundary element,  $g_{123,\emptyset}$ . For the objective function 2 bi-capacities in Table V, we can observe some elements that help push the background to 0 (Recall that Objective 2 pushes the non-target class to a neutral label of 0). For example, elements  $g_{1,\emptyset}$ ,  $g_{2,\emptyset}$ , and  $g_{12,\emptyset}$ , which are the weights in favor of sources 1 and 2, are closer to 0. This makes sense, as those sources show high confidence in background elements. This is in contrast to the previous bi-capacities, in which there is an inversion effect.

The normalized fuzzy measures learned by CI-QP, MICI, and MICI-BFM are shown in Table VI. The measures learned by these methods are similar as they all rely primarily on the element  $g_{123}$ , while the other measure element values are close to 0. This suggests that, unlike bi-capacities, normal fuzzy measures are unable to properly take advantage of the negative or background information in the sources, particularly from sources 1 and 2 in this experiment.

To quantitatively compare the different fusion methods, the Receiver Operating Characteristic (ROC) curve's Area Under Curve (AUC) and Root Mean Square Error (RMSE) metrics are computed between the result and the ground truth shown in Fig. 6a. A higher AUC signifies better target detection, while

TABLE V: Pedestrian detection bi-capacity learned with objective function 2. Bolded elements are used in the Bi-MICHI.

		Pedestrian Detection Bi-capacity (Obj. 2), $g_{A,B}$								
		$\emptyset$	1	2	12	3	13	23	123	
A	B	0	0	<b>-0.26</b>	<b>-0.40</b>	-0.59	<b>-0.68</b>	-1.00	<b>-0.69</b>	-1.00
		1	<b>0.11</b>	<b>0.00</b>		<b>0.09</b>		<b>0.00</b>		
		2	<b>0.07</b>	<b>-0.09</b>		-0.57	<b>-0.98</b>			
		12	<b>0.19</b>			<b>0.15</b>				
		3	<b>0.63</b>	0.61	<b>0.02</b>	-0.30				
		13	<b>0.79</b>		<b>0.02</b>					
		23	<b>1.00</b>		<b>1.00</b>					
		123	<b>1.00</b>							

TABLE VI: Measures learned by comparison Choquet integral methods (MICI, MICI-BFM, CI-QP).

Method	$g_1$	$g_2$	$g_3$	$g_{12}$	$g_{13}$	$g_{23}$	$g_{123}$
MICI	0.00	0.00	0.04	0.00	0.04	0.40	1
MICI-BFM	0	0	0	0	0	0	1
CI-QP	0	0	0	0	0	0.36	1

a lower RMSE signifies that the result is closer to the ground truth. Mask R-CNN, CI-QP, MICI, MICI-BFM, and our Bi-MICHI with objective function 2 produce confidence maps that are bounded from  $[0, 1]$ . These methods were normalized from  $[-1, 1]$  before the metrics were computed to offer a consistent comparison between all methods.

Table VII shows the results of these metrics. Interestingly, source 3 from the thermal sensor after gamma correction, along with weighted mean (which closely matches source 3), obtained a very high AUC for pedestrian detection. This is because the thermal sensor is able to sense the heat signature from humans even under low-light conditions. However, all input sources (including source 3) have a high RMSE value, which means all of them contain a lot of noise and false positives, especially in the non-pedestrian areas. Among the fusion methods, KNN and SVM performed really well on AUC, but both of these methods require pixel-level training labels and are highly sensitive to parameter choices. The Mask R-CNN method produced very low RMSE, but based

on the visual result (Fig. 7j), Mask R-CNN hardly detected any pedestrians and simply classified almost everything to background (thus, a low AUC). Among the Choquet integral-based methods (bottom five rows), our proposed Bi-MICHI achieves superior performance with higher AUC and lower RMSE compared to other Choquet integral-based methods using normalized fuzzy measures. This also shows that the bi-capacities proposed in this work are effective for detecting target classes (humans) while removing background noise, given input sensor sources that carry complementary information. Additionally, our proposed Bi-MICHI are trained with only bag-level labels, which allows the fusion process to take label uncertainty into consideration.

### C. Discussions

This paper tests two objective functions with the proposed Bi-MICHI algorithm. The first does not enforce the  $g_{\emptyset, \emptyset} = 0$  bound when constructing the bi-capacities, while the second does. In our experiments, we find that they are both useful for different applications. Objective function 2 works well if separate sources highlight different parts of the target. For example, in the synthetic “U”“M” experiment that uses objective function 2, the “U” is detected strongly positively in source 1 and the “M” is detected strongly negatively in source 2. Instead of trying to push both to 1 like objective function 1 would do, each letter can be pushed to the closest pole (+1 for “U” and -1 for “M”).

On the other hand, objective function 1 can be useful in situations where multiple sources detect the target, but on different ends of the  $[-1, 1]$  scale. For example, in the pedestrian detection experiment, pedestrians appear negatively in source 2 and positively in source 3. The “inversion” created by the objective function 2 allows the background to be pushed to -1 while still encouraging the pedestrians to label +1.

The use of bi-capacities and the Choquet Integral offers an explainable representation of the interactions between the sources to be fused. By examining the element values of the bi-capacities, users can interpret what role a sensor source plays in the final fusion result. Other machine learning techniques such as KNN, SVM, and Mask R-CNN, while effective in some cases, could be considered a “black box” in the sense that they provide predictions without a clear, interpretable reason a decision was made. The Choquet Integral, on the other hand, provides a clear, analytical model for complex, nonlinear fusion, and the bi-capacities provide an interpretable and explainable representation of how fusion sources are weighed against each other.

## VI. CONCLUSION

This paper presents Bi-MICHI, a novel Choquet integral-based method for explainable and interpretable data fusion with uncertain labels using bi-capacities. Through both synthetic and real-world experiments, this framework showed promising results in its ability to utilize negative and complementary information from input sensor sources for effective multi-sensor fusion.

Bi-MICHI aims to provide a general framework of data fusion that can be applied to various modalities and applications. To further explore the effectiveness of this method, additional experiments can be conducted applying Bi-MICHI to other fusion applications and sensor modalities. For example, this method can be used for multi-temporal and multi-view fusion [5], where a sequence of images can be fused for target detection and classification.

## REFERENCES

- [1] Michel Grabisch, "The application of fuzzy integrals in multicriteria decision making," *European Journal of Operational Research*, vol. 89, no. 3, pp. 445–456, 1996.
- [2] Paul Gader, Andres Mendez-Vasquez, Kenneth Chamberlin, Jeremy Bolton, and Alina Zare, "Multi-sensor and algorithm fusion with the Choquet integral: applications to landmine detection," in *IEEE Int. Symp. Geosci. Remote Sens.*, Sept. 2004, vol. 3, pp. 1605–1608.
- [3] Siva Krishna Kakula, Anthony J Pinar, Timothy C Havens, and Derek T Anderson, "Online sequential learning of fuzzy measures for choquet integral fusion," in *IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE)*, 2021, pp. 1–6.
- [4] Maria Skublewska-Paszkowska, Paweł Karczmarek, Paweł Powroznik, Edyta Lukasik, Jakub Smolka, and Michał Dolecki, "Aggregation of tennis multivariate time-series using the choquet integral and its generalizations," in *IEEE Int. Conf. Fuzzy Systems (FUZZ)*, 2023, pp. 1–6.
- [5] Brendan J Alvey and Derek T Anderson, "Geometrically motivated fuzzy measures and choquet integral for multi-look fusion," in *IEEE Int. Conf. Fuzzy Systems (FUZZ)*, 2023, pp. 1–8.
- [6] Derek T Anderson, Paul Elmore, Fred Petry, and Timothy C Havens, "Fuzzy choquet integration of homogeneous possibility and probability distributions," *Information Sciences*, vol. 363, pp. 24–39, 2016.
- [7] Xiaoxiao Du, Alina Zare, James M Keller, and Derek T Anderson, "Multiple instance choquet integral for classifier fusion," in *IEEE Congr. Evol. Comput.*, 2016, pp. 1054–1061.
- [8] Xiaoxiao Du and Alina Zare, "Multiple instance choquet integral classifier fusion and regression for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2741–2753, 2019.
- [9] Yasuo Narukawa and Vicenç Torra, "An score index for hesitant fuzzy sets based on the choquet integral," in *IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE)*, 2021, pp. 1–5.
- [10] Xiaohong Zhang, Jingqian Wang, Jianming Zhan, and Jianhua Dai, "Fuzzy measures and choquet integrals based on fuzzy covering rough sets," *IEEE Trans. Fuzzy Systems*, vol. 30, no. 7, pp. 2360–2374, 2021.
- [11] Michel Grabisch and Christophe Labreuche, "Bi-capacities – part i: definition, möbius transform and interaction," 2007.
- [12] Michel Grabisch and Christophe Labreuche, "Bi-capacities-II: the choquet integral," *Fuzzy sets and systems*, vol. 151, no. 2, pp. 237–259, 2005.
- [13] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
- [14] Dip Ghosh and Sanghamitra Bandyopadhyay, "A fuzzy citation-knn algorithm for multiple instance learning," in *IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE)*, 2015, pp. 1–8.
- [15] Jabbar Abbas, "Bipolar choquet integral of fuzzy events," in *IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*, 2014, pp. 116–123.
- [16] Jabbar Abbas, "The balancing bipolar choquet integrals," *International Journal of Innovative Computing, Information and Control*, vol. 17, no. 3, pp. 949–957, 2021.
- [17] Salvatore Corrente, José Rui Figueira, and Salvatore Greco, "Interaction of criteria and robust ordinal regression in bi-polar promethee methods," in *Advances in Computational Intelligence*, Salvatore Greco, Bernadette Bouchon-Meunier, Giulianella Coletti, Mario Fedrizzi, Benedetto Matarazzo, and Ronald R. Yager, Eds., Berlin, Heidelberg, 2012, pp. 469–479, Springer Berlin Heidelberg.
- [18] Ling Zhang, Yan Xu, Chung-Hsing Yeh, Le He, and De-Qun Zhou, "Bi-topsis: A new multicriteria decision making method for interrelated criteria with bipolar measurement," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 12, pp. 3272–3283, 2017.
- [19] Ling Zhang, Fandi Wang, Yan Xu, Chung-Hsing Yeh, and Peng Zhou, "Evaluating and selecting renewable energy sources for a microgrid: A bi-capacity-based multi-criteria decision making approach," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 921–931, 2021.
- [20] Fang Peng, Ling Zhang, and Yun Shi, "A new method for the dynamic comprehensive evaluation of regional eco-efficiency: Evidence from jiangsu province," *IOP Conference Series: Earth and Environmental Science*, vol. 804, pp. 042040, 07 2021.
- [21] Andres Mendez-Vazquez, Paul D. Gader, James M. Keller, and Kenneth Chamberlin, "Minimum classification error training for choquet integrals with applications to landmine detection," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 225–238, 2008.
- [22] Qinghua Wang, Canliang Zheng, Hongtao Yu, and Donghua Deng, "Integration of heterogeneous classifiers based on choquet fuzzy integral," in *7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Aug. 2015, vol. 1, pp. 543–547.
- [23] K. Hirota, H. A. Vu, P. Q. Le, C. Faticah, Z. Liu, Y. Tang, M. L. Tangel, Z. Mu, B. Sun, F. Yan, D. Masano, O. Thet, M. Yamaguchi, F. Dong, and Y. Yamazaki, "Multimodal gesture recognition based on choquet integral," in *IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE)*, 2011, pp. 772–776.
- [24] Wang Guan, Rong Yang, Yun Wang, Zijie Xie, and Junyu Niu, "Classifiers fusion of pedestrian detection based on choquet integral," in *IEEE International Conference on Real-time Computing and Robotics*, 2020, pp. 50–55.
- [25] Gwenolé Quellec, Guy Cazuguel, Béatrice Cochener, and Matthieu Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 213–234, 2017.
- [26] Seniya Esen Yuksel, Jeremy Bolton, and Paul Gader, "Multiple-instance hidden markov models with applications to landmine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6766–6775, 2015.
- [27] Achut Manandhar, Peter A. Torrione, Leslie M. Collins, and Kenneth D. Morton, "Multiple-instance hidden markov model for gpr-based landmine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1737–1745, 2015.
- [28] Andrew Karem and Hichem Frigui, "A multiple instance learning approach for landmine detection using ground penetrating radar," in *IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 878–881.
- [29] Oren Z. Kraus, Jimmy Lei Ba, and Brendan J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 06 2016.
- [30] J. Troy Cobb, Xiaoxiao Du, Alina Zare, and Matthew Emigh, "Multiple-instance learning-based sonar image classification," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*, Steven S. Bishop and Jason C. Isaacs, Eds. International Society for Optics and Photonics, 2017, vol. 10182, p. 101820H, SPIE.
- [31] Alina Zare, Changzhe Jiao, and Taylor Glenn, "Discriminative multiple instance hyperspectral target characterization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2342–2354, 2018.
- [32] Toshiaki Murofushi, Michio Sugeno, et al., "Fuzzy measures and fuzzy integrals," *Fuzzy measures and integrals: theory and applications*, vol. 2000, pp. 3–41, 2000.
- [33] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein, *Introduction to algorithms*, MIT press, 2022.
- [34] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [35] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baselines," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [36] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [37] Xiaoxiao Du, Alina Zare, and Derek T Anderson, "Multiple instance choquet integral with binary fuzzy measures for remote sensing classifier fusion with imprecise labels," in *IEEE Symp. Comput. Intell.*, 2019, pp. 1154–1162.