Contextual fusion of classifiers under variable atmospheric conditions for coastal surveillance

Océane Tran
CS Group
University of Toulon
La Garde, France
oceane.tran@csgroup.eu

Anne-Laure Jousselme
CS LAB / BL Marine
CS Group
La Garde, France
anne-laure.jousselme@csgroup.eu

Gilles Tedeschi
MIO
University of Toulon
La Garde, France
gilles.tedeschi@univ-tln.fr

Lionel Gardenal
CS LAB / BL Marine
CS Group
La Garde, France
lionel.gardenal@csgroup.eu

***Abstract*—Monitoring maritime areas is a challenging task which generally benefits from sensing systems such as radars or cameras, with different electromagnetic or optical capabilities. Infrared (IR) cameras in particular allow night and day surveillance, but their image is affected by maritime atmospheric conditions, e.g., the aerosol concentration. Further classification process thus inherits this image degradation resulting in possibly poor target classification results. To overcome this issue, we propose in this paper a contextual classifiers fusion system where two neural networks are trained into two environmental contexts and further combined with Bayesian reasoning. Individual classifier's reliability is considered, enabling to balance between the two classifiers depending on the uncertain context of use. Additionally, we apply an imprecise decision rule for a greater flexibility allowing a compromise between two criteria of accuracy and specificity. Results are obtained on simulated classifiers' outputs as well as on a synthetic dataset of IR images degraded with atmospheric context. It is shown that the proposed approach allows the possibility to increase accuracy while foregoing specificity and adds explainability.**

***Index Terms*—Maritime surveillance; image classification; Bayesian network; marine aerosols; imprecise classifier.**

I. Introduction

Control of littoral activities and maritime surveillance in general require sensing devices covering specific areas of interest, with different ranges, for early detection of anomalous events at sea. Manual coastal surveillance work is challenging for operators because they might miss a key information in a lot of situations. Long hours of work can diminish alertness and focus. The surveillance zone can be crowded and thus suspicious ships can hide amid the mass. Moreover, atmospheric conditions can decrease the visibility range and negatively impact the target recognition task. Visibility also decreases during night time. In such situations, operators would benefit from semi-automated surveillance tools to support their daily tasks, improving the detection range, the recognition and identification of vessels and enabling them to focus on events of interest.

Infrared images are used for vessel recognition tasks specially during night episodes. In this waveband, machine learning algorithms are sensitive to environmental variations during rain or fog episodes but also in presence of maritime aerosols created by sea sprays. The impact of the environment on ship classification is for instance considered in [1] where a method based on Convolutional Neural Network (CNN) and data augmentation with weather challenged data was proposed.

To address this problem of uncertain and possibly variable atmospheric conditions, we proposed a classification system composed of (1) two contextually trained neural networks, (2) combined with a Bayesian approach considering the reliability of the individual classifiers and (3) with an imprecise decision (labeling) function.

In Section II, we provide some background on classification in uncertain context, describing the environmental parameters and phenomena impacting the infrared image quality. The classification problem is also formalized, with an emphasis on imprecise decision. The proposed classifier system is described in Section III, as a contextually trained two-classifier system, with a Bayesian combination of outputs. Section IV is dedicated to results obtained on both simulated classifier's output and modified synthetic IR images of ships. We finally conclude in Section V and open on future work.

II. Classification in uncertain context

Coastal surveillance relies on the ability of the different authorities to detect and identify any suspicious or anomalous event at sea which can cause safety or security issues and would require an early intervention. Vessel Traffic Services (VTS) are shore-side systems which provide situational awareness to operators by means of sensing and communication devices, gathering and analysing information about the maritime traffic or meteorological hazard. While the most common sensors are radar systems, Automatic Identification Systems (AIS) or radio direction finders, cameras are often used to complement the maritime picture. Environmental conditions such as fog, wind, sun light reflection or marine aerosols [2], [3] have an impact on the signal, and thus on further processing until the outcome sent to the user. Visible cameras use wavelengths similar to the spectrum that the human eye perceives. Working on visible data is challenging because of the dynamic sea background, varying angles of view, variation in lighting conditions and a lot of different scale ratios in the same image. Although some methods exist to handle these aspects, they do not or barely consider severe weather conditions [4]–[6]. Infrared cameras instead provide data which is much more invariant than visible sensor to

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00171

atmospheric irradiance such as solar glint and offer a wide range of use from daylight to night conditions. In this paper, we focus on the problem of ship recognition with degraded IR images by environmental conditions.

A. From observation to user

Although fog, glint and low irradiance are conditions handled by infrared sensing, this waveband is sensitive to attenuation due to environmental conditions variations. In conjunction with humidity, the atmospheric visibility is affected by molecular and aerosol attenuation[1]. In extreme conditions, high wind and high sea state drastically increase the marine aerosol concentration which degrades thus the quality of the IR image captured by the camera. For instance, the visibility loss due to sea spray aerosols can reach up to 0.2 km^{-1} [7]. We do not take into account other aerosols like dust or pollution.

Vessel emits and reflects electromagnetic signal. The scene environment adds noise and distorts vessel's signal (diffusion, absorption, attenuation from molecules in atmosphere and aerosols from sea spray and dust, adverse weather conditions [1]). Then the sensor affects the radiance emitted by the ship because of its resolution, spectral bandwidth, distance, pitch and camera response function. The sensor output is an image with specific contrast, saturation and sharpness which can be modified by software in acquisition process. The classification process aims at labeling the image of a vessel and depends on classification method and training data. The labeling process finally outputs the decision for the operator and can be prone to errors, lack of explanation or not adapted to operational needs.

In this paper, we define *context* as the conditions under which the classifier is to be used. We distinguish between environmental context (for instance, sea spray aerosols visibility loss in this case) and operational context (e.g., user needs). For instance, the operator may not be interested in knowing the precise type of a vessel, but only in distinguishing between military and civilian vessels, while a finer-grained set of classes may be required later.

In the following, we will assume that the environmental conditions are gathered under a variable of visibility which impacts the classifiers' reliability in their ability to provide correct classifications (see Section III-B).

B. Imprecise classification

Imprecise classification allows selecting several class as part of the decision. Different methods have been proposed framed in different uncertainty theories such as belief functions (e.g., [8], [9]), possibility (e.g., [10]) or credal sets (e.g., [11]).

Formally, we consider a set of N classes $\Omega = \{\omega_1, ..., \omega_N\}$ and a set of M samples $\mathcal{X} = \{x_1, ..., x_M\}$. An *imprecise* classification function Ψ assigns one or several classes to a sample, such that $\Psi(x_m) \subseteq \Omega$ is the classifier's estimate of a subset of classes possibly containing the true classes. The true class of sample x_m is denoted by ω_m^*. This function Ψ can be seen as a composition of two sub-functions s and l, where $s : \mathcal{X} \to \mathbb{R}^N$ and $l : \mathbb{R}^N \to 2^\Omega$, with 2^Ω denoting the power set of Ω. s is the function which assigns scores to a sample, and l is the decision function. Precise classification consists in choosing one class (usually the class with the highest score) and is a special case of imprecise classification. Generally, the minimization of expected loss is used ([11], [12]) with a loss function as a compound measure of accuracy and specificity (linked to the number of classes output measures).

In the following, we will exploit the flexibility of imprecise classification to adapt the classifier's output to different contexts of use.

[1]The atmospheric visibility or meteorological range is defined quantitatively, eliminating the subjective nature of the observer

III. Contextual Bayesian reasoning for classifiers combination

We describe in this section the proposed approach as a classifier system involving two contextually trained classifiers combined with a Bayesian approach and a proper reliability model [13].

A. Classifier system with contextual training

The proposed classifier system is depicted in Fig. 1. The process is made of three steps: (1) the sample x_m is preprocessed by two classifiers trained previously on different contexts; (2) their attributed scores are combined with a classification fusion module based on Bayesian reasoning with evidence on context (see Section III-B); (3) the posterior probability is then handled by the imprecise decision function l_α (see Section III-C).

A classifier is linked to a context through its training data and defined by its subset of fixed parameters. For instance, a low visibility image dataset contains images with low contrast. In this paper, we consider two contexts of use as defined by the meteorological conditions impacting the visibility, and denoted as low and medium visibility, characterized by a 2 km and 10 km visibility respectively. Two classifiers are trained in their respective context, medium visibility for classifier 1 and low visibility for classifier 2. The reliability of the classifier depends on the context, as captured by the confusion matrices during the training phase.

B. Bayesian contextual reasoning

The Bayesian contextual reasoning for classifier's fusion follows the general evidential model proposed in [14], [15] and is displayed as Bayesian networks in Fig. 2.

We denote by X the random variable (and corresponding node of the network) the true type of the ship, by C_1 and C_2 the estimations from classifier 1 and 2 respectively, by R_1 and R_2 their corresponding reliability variables and by V the contextual visibility variable. Variables X, C_1 and C_2 have $\Omega = \{\mathrm{Corvette; Frigate; Destroyer; Ferry}\}$ as universe of discourse, R_1 and R_2 have two possible states $\{\mathrm{Reliable; Not\ reliable}\}$, while V can take values within $\{\mathrm{Low; Medium}\}$. The prior probability on X is assumed to be uniform. The conditional probabilities connecting variables X, C_1 and C_2 are excerpt from the classifiers' confusion matrices on training datasets.

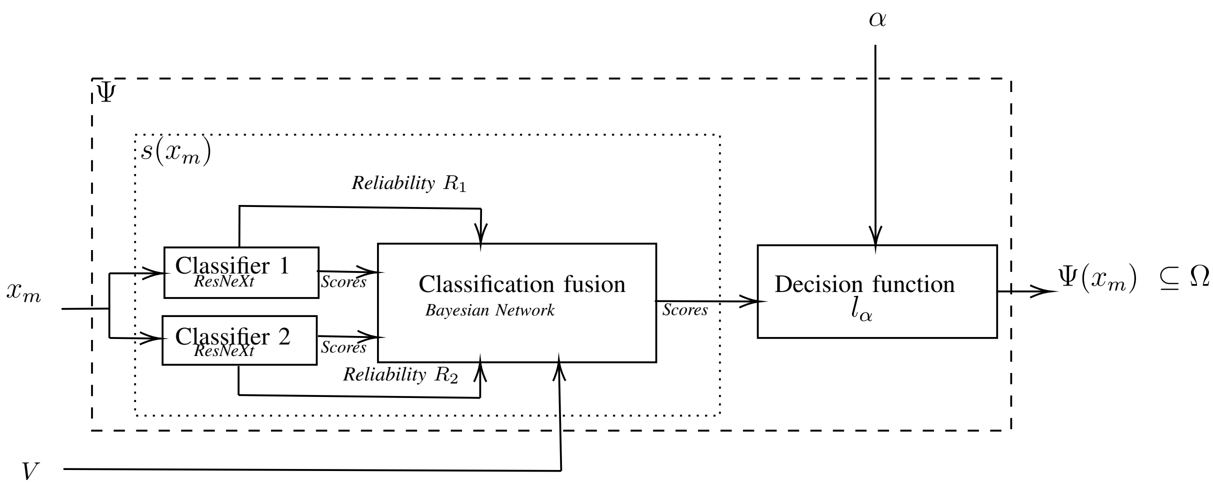
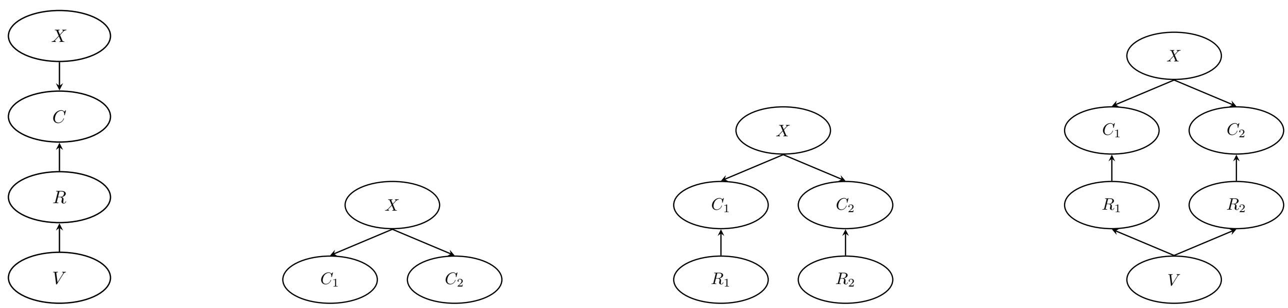


Fig. 1: Contextual fusion for cautious classification



(a) C1-3 (b) C2-1 (c) C2-2 (d) C2-3

Fig. 2: Four contextual classification architectures as Bayesian networks

The behavior of the classifier depends on its reliability R_i, $i \in \{1;2\}$. This reliability comes from the classifier context of use which itself can have two states, either medium or low visibility. The model implemented considers unreliable classifiers as randomizers [13]. Finally, classifier outputs are processed as soft evidence on variables C_1 and C_2. The originality of this model is to consider variable reliability depending on context.

The architectures are named Ca-b, where a is the number of classifiers and b corresponds to the step of contextual reliability involved. As such, C1-3 is the architecture with one classifier C_1 and the reliability R depending on the contextual visibility variable V. In the same way, C2-1, C2-2 and C2-3 are architectures with two classifiers, the later being the more complete one. Their performances will be compared in Section IV-A.

C. Variable thresholds for imprecise classification

We consider a new cautious way for selecting imprecise classes. Instead of considering scores themselves, we consider the relation and similarity between scores and define the following decision function $l_\alpha : \mathbb{R}^N \to 2^\Omega$ such that:

$$l_\alpha(s(x_m)) = \{\omega_j \in \Omega | s_j(x_m) \geq \alpha \max((s(x_m))\} \quad (1)$$

where $s_j(x_m)$ is the score associated with the class ω_j. The parameter $\alpha \in [0,1]$ dictates how scores should be alike to be considered.

D. Performance measures

Performances of the classifier systems are compared by means of a pair of measures which quantify respectively notions of accuracy and specificity. The accuracy is the measure proposed in [8]:

$$acc = \frac{1}{M} \cdot \sum_{m=1}^{M} \delta(\omega_m^*, \Psi(x_m)) \quad (2)$$

with

$$\delta(\omega_m^*, \Psi(x_m)) = \begin{cases} 1 & \text{if } \omega_m^* \subseteq \Psi(x_m) \\ 0 & \text{else} \end{cases} \quad (3)$$

while we define a specificity measure as:

$$spe = 1 - \frac{1}{M \cdot acc} \cdot \sum_{m=1}^{M} \delta(\omega_m^*, \Psi(x_m)) \frac{\log |\Psi(x_m)|}{\log N} \quad (4)$$

where $|.|$ denotes the cardinality operator. Specificity in this paper is based on a normliased version of Hartley entropy, where only correctly classified samples are considered. It should not be confused with true negative rate like in confusion matrix. The acc measure quantifies how much the classifier system's outputs are correct ($acc = 1$ if it never makes mistakes), and the spe measure quantifies how much the classifier's outputs are specific ($spe = 1$ if all outputs are singletons).

IV. Results

We demonstrate below the benefits of the proposed approach for the classification of vessels with uncertain context, on simulated outputs of imprecise classifiers in Section IV-A, and on synthetic IR images of ships in Section IV-B.

A. On simulated imprecise classifiers

In order to test our approach with full control on parameters, we designed a simulator of imprecise classification outputs. The parameters are the number of classes N, the number of samples per class, the global accuracy (mean and variance), the ratio of non-specific outputs (mean and variance), the maximal number of classes in case of a non-specific output and a likeness criterion for non-specific outputs. The simulator produces an imprecise confusion matrix in percentage before converting it into numbers of samples and producing a list of imprecise decisions and corresponding scores. Fig. 3 displays the results obtained with simulated imprecise classification results, according to the four architectures of Fig. 2. The curves specificity/accuracy are obtained for different values of α (1). Performances of the two basic simulated classifiers are displayed with dashed lines.

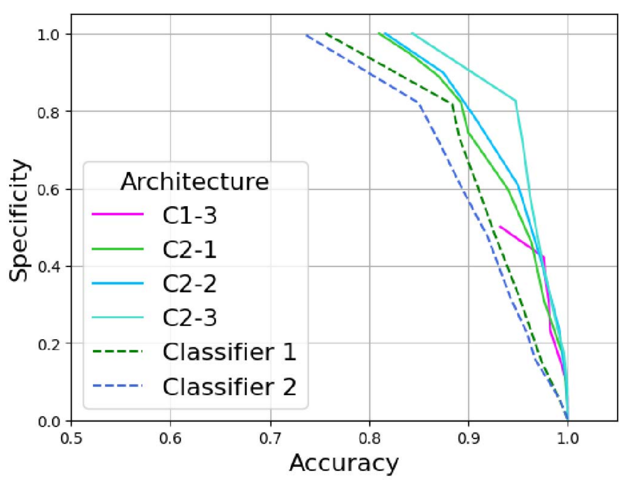


Fig. 3: Performances in accuracy and specificity of our classification method with simulated imprecise classifiers

We can observe in Fig. 3 that all systems have a better performance than the two basic classifiers. For a given specificity, accuracy is systematically improved.

The effects of combination can be observed by comparing C1-3 and C2-3. Although accuracy values are similar for low specificity values, the system with two classifiers C2-3 allows more modulation on specificity, which ranges from 0 to 1 instead of 0 to 0.5 for C1-3. The reason behind this phenomenon lies in the statistical nature of the specificity measure. Indeed, C1-3 decision cardinality is maximal for 50% of the samples because the system is reliable on only one context while C2-3 is reliable on both contexts and enables full specificity.

The effects of adding the reliability variable in the model are exposed by comparing C2-1 and C2-2. We observe that performances are really close to each others with an advantage for C2-2. Indeed, the reliability variable mildly increases the accuracy when the specificity decreases.

Finally, the effects of adding context to reliability to form **contextual reliability** can be observed by comparing C2-2 and C2-3. This comparison shows that with two classifiers, for a given specificity contextual reliability increases the accuracy (up to 28% when spe = 1). Also, the system makes less mistakes at maximal specificity. And we could obtain a good compromise of an accuracy of 95% at the expense of a specificity of circa 81%.

B. On synthetic IR images with real classifier

1) Data: We used the IRShips database gathering, synthetic IR of 9 military ships and one ferry, grouped into 4 classes of Corvette, Frigate, Destroyer and Ferry [16]. Each image is described by a series of parameters including ship name, thermal appearances, ranges, bearings and elevations. The dataset has then been modified following four steps: (1) 1024 images per ship were selected from the original dataset while keeping maximum variability in terms of bearing and range, (2) images were then cropped and centered around ships resized to the same dimension, (3) the dataset was divided into two subsets and (4) one context was applied on each subset.

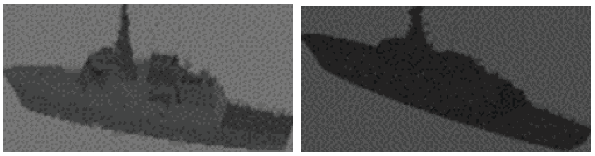
For Step (4), images were thus modified to represent two contexts with varying visibility according to Beer-Lambert law and Kim model for optical attenuation [17]:

$$I = I_0 \cdot e^{-\beta \cdot d} + R \tag{5}$$

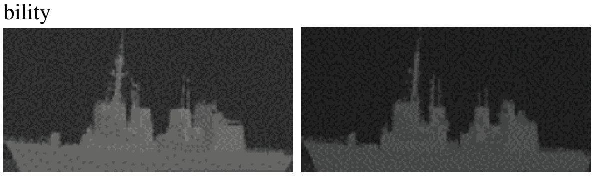
where I is the modified image, I_0 the original image,d the distance of the ship from the camera (in km) and R a normal random matrix with the same size as the image, $R \sim \mathcal{MN}_{n,p}(\mu, \sigma)$ with $\mu = 2$ and $\sigma = 2$. The atmospheric attenuation coefficient β is used to simulate an atmospheric veil due to sea spray. We are using $\beta = 1.98$ for low visibility and $\beta = 0.16$ for medium visibility. Atmospheric turbulence is not taken into account in analysis.

The resulting contextual datasubsets are named $ctxt1$ and $ctxt2$ for context 1 and context 2 respectively. Examples of resulting images are presented in Fig. 4.

The ResNeXt50 algorithm [18] was selected due to its high accuracy and for its easy set up. The models are pre-trained



(a) Corvette with medium visibility (b) Corvette with low visibility



(c) Destroyer with medium visibility (d) Destroyer with low visibility

Fig. 4: Veiled ship images from the IRships database at medium and low visibility with Eq. (5)

on ImageNet[2] by PyTorch[3] and used for transfer-learning on our contextual datasets. Both subdatasets were concatenated into the dataset $ctxt0$ that was used to create a reference classifier with the same parameters. Each classifier is named after the context of its training dataset: Classifier 1 was trained on $ctxt1$, Classifier 2 was trained on $ctxt2$ and Classifier 0 was trained on $ctxt0$. The three classifiers were trained with the same hyper-parameters, obtaining the average performances of 89,97%, 79,60%, and 96,33% respectively on a 6-fold cross-validation for error estimation. As expected, Classifier 0 performs better than the two others because it was trained with twice the images number.

The general dataset was sliced to build separate training and test datasets with 1/6 ratio.

2) Performances: Among all the architectures we focused on C2-3 which gives the most interesting results. In Figure 5a, performances are shown for a variable context, thus mixing observations with both low and medium visibility. Although the reference classifier (red dashed line) produces the best accuracy for the maximum specificity, it remains limited in decreasing specificity, and thus increasing accuracy. In practice, even if the user is willing to lose some specificity for gaining accuracy Classifier 0 would not offer that flexibility. C2-3 instead enables a gain of 10% (to circa 95%) in accuracy for a specificity of 80%.

Figures 5b and 5c show that this result is not always true and depends on the context. When operating conditions with medium visibility (Fig. 5b), Classifier 0 performs better than Classifier 1 (trained itself only with images with medium visibility). This shows that adding data with low visibility to training data increases slightly the accuracy. When operating conditions with low visibility (Fig. 5c), Classifier 2 is more accurate than Classifier 0 while C2-3 allows more accuracy than Classifier 2 when lowering the specificity.

[2] www.image-net.org/index.php
[3] www.pytorch.org

Those results convey that under variable atmospheric conditions, our classification system C2-3 is at least as good as the reference classifier 0 but allows to make less mistakes at the expense of some loss in specificity.

V. Conclusions and future work

In order to automatically adapt to varying atmospheric conditions while performing automatic target recognition tasks, we proposed in this paper an hybrid classification system built from two pre-trained CNNs under specific visibility conditions combined by a Bayesian reasoning, which enables to contextually consider the classifiers' reliability. Furthermore, we propose an original imprecise labeling rule enabling classifiers to output non-specific sets of classes for a greater flexibility in improving accuracy. The efficiency of the proposed approach was demonstrated on simulated classifiers' outputs as well as on synthetic IR images of ships. The results show that with medium visibility, a classifier trained on more data gives better accuracy than one specialized, trained on that specific context. But with low visibility, a specialized classifier and our C2-3 method enables more flexibility to increase accuracy with possible specificity loss. This is an important practical result, as the hybrid classifier system C2-3 enables to adapt to possible varying user needs. In future work, we will extend to other frameworks for uncertainty reasoning, in particular belief functions, while the C2-3 classifier will be validated *in situ* with testing on the French Mediterranean coastline.

Acknowledgment

This PhD thesis work is supported by the French DGA-AID (Direction Générale de l'Armement - Agence de l'Innovation de la Défense), supervised by Frédéric Livernet, DGA TN (Techniques Navales), as part of a collaboration agreement between CS Group, U. of Toulon (MIO) and the NATO STO CMRE. We thank Yvonick Hurtaud from DGA MI for his insightful comments.

References

[1] X. Nie, M. Yang, and R. W. Liu, "Deep neural network-based robust ship detection under different weather conditions," in *22nd IEEE Int. Conference on Intelligent Transportation*, (Auckland, New Zealand), 2019.

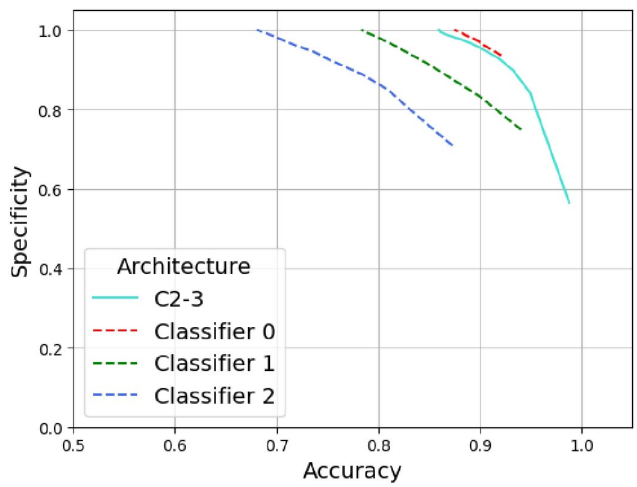
[2] G. Tedeschi, A. M. van Eijk, J. Piazzola, and J. T. Kusmierczyk-Michulec, "Influence of the surf zone on the marine aerosol concentration in a coastal area," *Boundary-Layer Meteorology*, vol. 163, pp. 327–350, 2017.

[3] A. Demoisson, G. Tedeschi, and J. Piazzola, "A model for the atmospheric transport of sea-salt particles in coastal areas," *Atmospheric Research*, vol. 132-133, pp. 144–153, 2013.

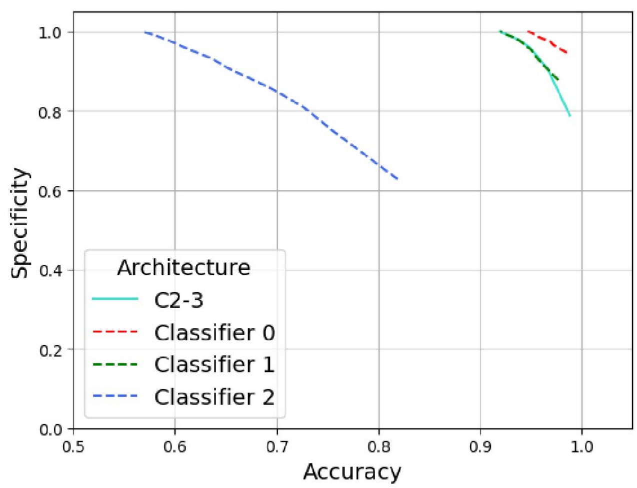
[4] F. Bousetouane and B. Morris, "Fast CNN surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios," in *13th IEEE Int. Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 242–248, Aug. 2016.

[5] M. Leclerc, R. Tharmarasa, M. C. Florea, A. Boury-Brisset, T. Kirubarajan, and N. Duclos-Hindié, "Ship Classification Using Deep Learning Techniques for Maritime Target Tracking," in *21st Int. Conference on Information Fusion (FUSION)*, pp. 737–744, July 2018.

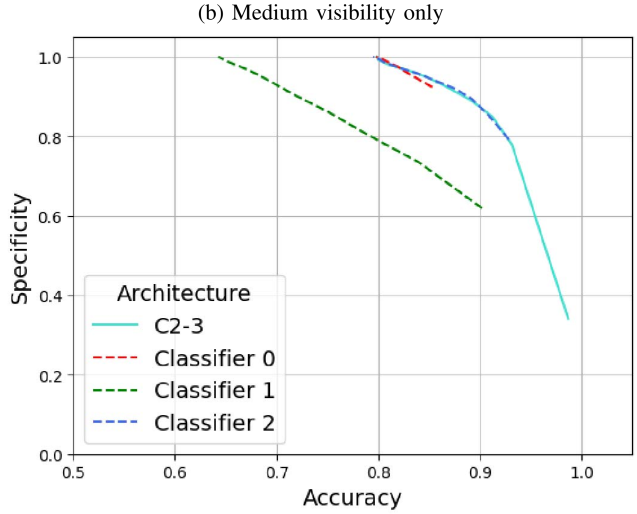
[6] C. Chen and H. Lang, "Multiple Features Learning for Ship Classification in Optical Imagery," *Multimedia Tools and Applications*, vol. 77, pp. 13363–13389, June 2018.



(a) Variable (mixed medium and low) visibility



(b) Medium visibility only



(c) Low visibility only

Fig. 5: C2-3 performances on testing datasets for variable, medium and low visibility

[7] Y. Su, Y. Han, H. Luo, Y. Zhang, S. Shao, and X. Xie, "Physical-Optical Properties of Marine Aerosols over the South China Sea: Shipboard Measurements and MERRA-2 Reanalysis," *Remote Sensing*, vol. 14, p. 2453, Jan. 2022.

[8] L. Jacquin, *Traitement de données incertaines : application au tri des matières plastiques*. These de doctorat, IMT Mines Alès, Mar. 2021.

[9] Z.-g. Liu, Q. Pan, and J. Dezert, "Classification of uncertain and imprecise data based on evidence theory," *Neurocomputing*, vol. 133, pp. 459–470, June 2014.

[10] V.-L. Nguyen, S. Destercke, M.-H. Masson, and E. Hüllermeier, "Reliable multi-class classification based on pairwise epistemic and aleatoric uncertainty," in *Proceedings of the 27th Int. Joint Conference on Artificial Intelligence*, IJCAI'18, (Stockholm, Sweden), pp. 5089–5095, AAAI Press, 2018.

[11] M. Zaffalon, G. Corani, and D. Mauá, "Evaluating credal classifiers by utility-discounted predictive accuracy," *Int. Journal of Approximate Reasoning*, vol. 53, pp. 1282–1301, Nov. 2012.

[12] J. J. del Coz Velasco, J. Díez Peláez, and A. Bahamonde Rionda, "Learning nondeterministic classifiers," *Journal of Machine Learning Research, 10*, vol. 10, pp. 2273–2293, 2009.

[13] L. Bovens and S. Hartmann, *Bayesian Epistemology*. Oxford University Press, 2003.

[14] P. Kowalski and A.-L. Jousselme, "Multi-source contextual reasoning for vessel behavourial analysis," Tech CMRE-FR-2020-002, NATO STO Centre for Maritime Research and Experimentation (CMRE), 2020.

[15] P. Kowalski and A.-L. Jousselme, "Context-awareness for information correction and reasoning in evidence theory," *International Journal of Approximate Reasoning*, Nov. 2022. In press.

[16] S. T. Westlake, T. N. Volonakis, J. Jackman, D. B. James, and A. Sherriff, "Deep learning for automatic target recognition with real and synthetic infrared maritime imagery," in *Artificial Intelligence and Machine Learning in Defense Applications II*, vol. 11543, pp. 41–53, SPIE, Sept. 2020.

[17] I. I. Kim, B. McArthur, and E. J. Korevaar, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," in *SPIE Optics East* (E. J. Korevaar, ed.), (Boston, MA), pp. 26–37, Feb. 2001.

[18] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, 07 2017.