



Domain-informed and neural-optimized belief assignments: A framework applied to cultural heritage

Sofiane Daimellah ^{a,*}, Sylvie Le Hégarat-Mascle ^a, Clotilde Boust ^b

^a Université Paris-Saclay, 9 Rue Joliot Curie, Gif-sur-Yvette, 91190, France

^b Centre de Recherche et de Restauration des Musées de France, 14 Quai François Mitterrand, Paris, 75001, France

ARTICLE INFO

Keywords:

Belief function theory
Mass function assignment
Neural network-based optimization
Cultural heritage
Pigment identification

ABSTRACT

Identifying pigments in Cultural Heritage artifacts is key to uncovering their origin and guiding conservation strategies. Although recent advances in non-invasive imaging have enabled the collection of rich multimodal data, existing methods often fall short in dealing with uncertain, ambiguous, or noisy information. This paper introduces a versatile fusion framework grounded in Belief Function Theory, combining domain-informed evidence modeling with neural optimization. Specifically, we propose a general strategy for assigning mass functions by leveraging expert knowledge encoded in parametric Evidence Mapping Functions, which are further refined through task-specific training using constrained neural networks. When applied to pigment classification, our method demonstrates robustness against source variability and class ambiguity. Experiments conducted on both synthetic and mock-up datasets validate its effectiveness and suggest promising potential for broader applications.

1. Introduction

Accurate pigment identification is essential for preserving Cultural Heritage (CH)¹ objects such as ancient paintings and photographs, as it provides information about their provenance and production period, while informing appropriate conservation strategies [57]. Traditionally, pigments have been analyzed using microsampling-based chemical analyses, which not only cause significant damage to precious artifacts but also rely heavily on human expertise [4,8]. Recent advances in non-invasive analysis have facilitated the acquisition of multimodal data, including X-Ray Fluorescence analysis (XRF), which determines the elemental composition of materials [9], and Reflectance Imaging Spectroscopy (RIS), which captures the reflection of light from materials across different wavelength ranges [61], all while preserving the integrity of artworks [38,63].

These advances have encouraged the CH community to investigate computational frameworks for various applications, focusing primarily on automatic pigment identification and unmixing [41]. Although these efforts have yielded promising results, conventional approaches that address pigment identification have some limitations depending on the modality used. For example, XRF-based methods struggle to distinguish pigments that share the same characteristic chemical elements, often grouping them into a single pigment class [33,65]. Furthermore, as a surface-sensitive technique, RIS may not provide information about pigments in the underlying layers, thereby limiting the reliability of RIS-based approaches in multilayered artworks [52,53]. To overcome these limitations, one

* Corresponding author.

E-mail address: sofiane.daimellah@universite-paris-saclay.fr (S. Daimellah).

¹ A list of all abbreviations is provided in Appendix A.

<https://doi.org/10.1016/j.ijar.2025.109534>

Received 29 April 2025; Received in revised form 4 July 2025; Accepted 22 July 2025

can leverage the partial yet complementary information from different modalities, as a multimodal approach is expected to resolve the ambiguities inherent in single modalities and enables more accurate pigment identification.

Multimodal systems represent, reason about, and fuse information from multiple data sources to accomplish a specific task in a comprehensive way that unimodal systems cannot achieve. By taking into account multiple sources of information, data fusion enables more reliable solutions in a wide range of applications; from autonomous navigation in mobile robotics [35,3] to medical diagnosis [42,55], through environmental applications such as precision agriculture [2]. Data fusion techniques range from deep learning approaches, among which multimodal learning (MML) is gaining popularity [68], to traditional Bayesian frameworks, such as Bayesian filtering, which sequentially fuse information by updating posterior distributions with new observations [64]. As an alternative to probabilistic methods, some frameworks allow one to distinguish between different types of partial knowledge (e.g., uncertainty, imprecision). Among those, Belief Function Theory (BFT), also known as Dempster-Shafer theory, provides a flexible framework for representing and combining uncertain or imprecise information, making it particularly suitable for applications dominated by ambiguous or incomplete data [56,12].

A few multimodal methods have been proposed to address some problems in the CH field. For example, Cai et al. introduced a structured protocol that progressively narrows down the set of remaining possible pigments [5]. However, this approach is mostly deterministic and does not account for the inherent variability within the data. Alternatively, Partial Least Squares Discriminant Analysis and Random Forest have been explored for historical pigment classification [10] and compound distribution mapping [25], respectively. While effective, these techniques lack the ability to incorporate prior knowledge for representing source information and may struggle when data is scarce or ambiguous. Using two hyperspectral imaging modalities (VIS-NIR and SWIR), along with MA-rFTIR, [6] demonstrates the potential of such data for discriminating pictorial materials. The authors use an ECOC-SVM (Error Correcting Output Code - Support Vector Machine) approach to handle the large number of classes, but the specific coding and decoding strategies are not reported, limiting reproducibility and comparison. Moreover, the method does not explicitly handle uncertainty, exploit domain-specific knowledge, or integrate recent developments in the use of neural networks. This underscores the need for more robust fusion frameworks capable of representing partial ignorance and resolving conflicting information to address the imprecise and uncertain nature of data in the CH domain without relying on extensive datasets.

To bridge these gaps, we propose a novel multimodal pigment identification framework grounded in BFT, whose central contribution lies in a systematic method for assigning mass functions. The proposed approach explicitly incorporates domain expertise in the initial definition of evidence through parametric Evidence Mapping Functions (EMFs), and further refines this representation through a neural optimization process formulated as a Constrained Neural Network (Constrained-NN). Therefore, our contribution is twofold: from an application perspective, we introduce a model for multimodal pigment identification capable of resolving ambiguities inherent to single-modality methods and overcoming the deterministic nature of existing multimodal protocols [5]; and from a methodological standpoint, we present a unified framework for mass function assignment that combines structured domain knowledge with data-driven learning, offering both interpretability and adaptability to real-world uncertainty.

The remainder of this paper is organized as follows. Section 2 reviews related work in the fields of CH and BFT. Section 3 introduces the proposed framework for mass function assignment in a general multimodal setting. Section 4 details its instantiation for the task of multimodal pigment identification. Experimental results are reported and analyzed in Section 5. Finally, Section 6 concludes the paper and outlines potential directions for future work.

2. Related works

This section begins with an overview of the main modalities used in CH and the existing pigment classification techniques in Section 2.1. The basic principles of BFT, including mass function representation, evidence combination, and decision-making, are then recalled in Section 2.2. Finally, Section 2 reviews the existing methods for mass function assignment, distinguishing between similarity-based and optimization-based approaches.

2.1. Cultural heritage

2.1.1. Modalities used in cultural heritage

For the purpose of artwork characterization, non-invasive techniques such as Hyperspectral Imaging, especially in reflectance mode [47], X-ray Powder Diffraction [25], and Raman Spectroscopy [59] provide insights into unique material properties, such as spectral, structural and molecular information. In this work, we used Optical Imaging and XRF spectroscopy due to their widespread availability in conservation laboratories, their ability to provide interpretable data, and their complementary nature. Optical imaging modalities include Visible (VIS), Ultraviolet False Color (UVFC) and Infrared False Color (IRFC) images: VIS images capture the reflection of red, green, and blue components in the electromagnetic spectrum (380-780 nm), forming three channels; UVFC and IRFC images are created by capturing light reflections in shifted spectral intervals: UVFC images combine UV-reflected (360-380 nm), green, and blue channels, while IRFC images combine IR-reflected (780-1000 nm), red, and green channels. XRF analysis yields spectra that provide valuable information about the chemical composition of pigments, from which the XRF maps that describe the distribution of chemical elements are derived using tools like Pymca [60].

2.1.2. Unimodal pigment classification

Unimodal methods rely on a single modality to identify individual pigments or groups of pigments that may remain ambiguous when the considered modality is unable to distinguish them. In terms of developed approaches, statistical methods and machine

learning algorithms such as Spectral Angle Mapper [11] and K-means Clustering [54] have been widely used, while NNs have recently gained popularity for their ability to model complex nonlinear patterns from spectral data. Multi-Layer Perceptrons (MLPs) [48] and Convolutional Neural Networks (CNNs) [40] are the two most commonly used models for unimodal pigment classification. For instance, a shallow MLP was trained on real samples [26], while two other deep MLP models [52,49] used synthetic reflectance spectra generated by Kubelka-Munk theory [67]. For most CH applications, 1D-CNNs are preferred over 2D-CNNs due to the spectral nature of the data, as spatial information is less relevant for material identification. For pigment classification, two approaches using 1D-CNNs have been proposed, which differ mainly in the type of data considered: in the first case [36], the model is trained on reflectance spectra of a real artwork with known ground truth, whereas in the second case, the model is pre-trained on a synthetic dataset of XRF spectra generated using the Fundamental Parameter method [58] and fine-tuned on labeled mock-ups [33,65].

As highlighted, single-modality approaches are inherently limited as they may not be able to distinguish certain groups of pigments (e.g., red lead and lead white in [33]). Moreover, training NNs on spectral data requires relatively large datasets that are not available within the CH community, and spectra generated from physical models such as Kubelka-Munk and Fundamental Parameter methods fail to fully simulate the complexity of physical phenomena or account for instrumental and environmental noise [41].

2.1.3. Multimodal pigment identification

Multimodal approaches combine complementary information from multiple sources to achieve a more comprehensive understanding of artworks and to address the ambiguities inherent in unimodal methods. Among the existing literature, the work in [5] proposes the only known general multimodal method for pigment identification. This protocol progressively narrows down the set of possible pigments through four stages:

1. **Production date:** The first stage uses the artwork's production date to exclude pigments that were not available during the relevant historical period.
2. **Optical imaging:** Optical imaging modalities are analyzed to differentiate pigments based on their characteristic color in each modality. For each pigment color, a decision tree is used where the nodes represent the remaining candidate pigments and the branches correspond to the responses from the imaging modalities.
3. **XRF analysis:** The third stage refines the list of candidate pigments by analyzing their chemical composition using XRF.
4. **Spectral Feature Analysis:** Finally, the reflectance spectrum of the sample is compared to a database of pigment spectra using similarity measures, such as Spectral Information Divergence [7], to resolve any remaining ambiguities regarding the present pigments.

Although the proposed method effectively addresses the ambiguities present in unimodal approaches, it is more of a roadmap that formalizes the protocol followed by a pigment identification expert, without actual automation, and without accounting for possible errors (e.g., due to noise) in the partial decisions. This contrast with advances in multimodal data uses in other fields, such as medical imaging and remote sensing, highlights a significant gap in existing multimodal frameworks for cultural heritage. It underscores the need for automated approaches that integrate multimodal data while addressing variability, single-modality ambiguities, and source reliability.

2.2. Basics in belief function theory

Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, referred to as the *Discernment Frame* (DF), be a set of mutually exclusive hypotheses representing all possible outcomes or states of a given problem. To account for uncertainty about subsets of hypotheses, we also consider the *powerset* of Ω , denoted 2^Ω , that is the set of all possible subsets of Ω . We represent evidence about an *uncertain variable* taking values in 2^Ω by a *mass function* such that:

$$m : 2^\Omega \rightarrow [0, 1], \quad \sum_{A \in 2^\Omega} m(A) = 1. \quad (1)$$

The mass $m(A)$ is interpreted as a measure of the belief that is assigned exactly to hypothesis A and to no more specific hypothesis. Each $A \in 2^\Omega$ where $m(A) > 0$ is called a *focal element* and $F_m = \{A \in 2^\Omega, m(A) > 0\}$ is defined as the set of focal elements of m . The mass, $m(\Omega)$, attributed to Ω represents the degree of *ignorance* which results from the vagueness of an evidence. Total ignorance is then represented by the *vacuous* belief function such that $m_v(\Omega) = 1$.

2.2.1. Belief and plausibility functions

A mass function, m , is in one-to-one relationship with several other BFT functions, among them are the *belief function* Bel and the *plausibility function* Pl , defined for all $A \subseteq \Omega$ as:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

$$Pl(A) = \sum_{\substack{B \subseteq \Omega \\ B \cap A \neq \emptyset}} m(B)$$

The quantity $Bel(A)$ is interpreted as the amount of evidence supporting A , while $Pl(A)$ represents the amount of evidence that is consistent with A , i.e. not supporting \bar{A} (the complement of A). The contour function of m is then defined for all $\omega \in \Omega$ as:

$$pl(\omega) = Pl(\{\omega\})$$

2.2.2. Reliability of evidence

In practice, information sources are not equally reliable. In BFT, the reliability of a source is taken into account using correction mechanisms, among which is the global discounting operation, which is the simplest. Formally, a mass function m with an associated degree of reliability $\alpha \in [0, 1]$ is transformed into a less committed mass function m^α such that:

$$\begin{cases} m^\alpha(A) = \alpha.m(A), & \text{if } A \in 2^\Omega \setminus \{\Omega\} \\ m^\alpha(\Omega) = \alpha.m(\Omega) + (1 - \alpha). \end{cases}$$

2.2.3. Coarsening and refinement

Evidence from a given source may be expressed at varying levels of granularity. In particular if there is a mapping Γ between two DFs Θ and Ω such that $\forall \theta_i \in \Theta, \Gamma(\theta_i) \subseteq \Omega$ and $\{\Gamma(\theta_i), i \in [1, |\Theta|]\}$ defines a partition of Ω , then Θ is said a *coarsened* DF of Ω and $\Gamma : \Theta \rightarrow 2^\Omega \setminus \{\emptyset\}$ is a *refinement* mapping. By extending it to 2^Θ as follows:

$$\forall A \in 2^\Theta, \Gamma(A) = \bigcup_{\theta_i \in A} \Gamma(\theta_i),$$

one can transfer the belief from coarse DF Θ to refined DF Ω as follows:

$$\forall A \in 2^\Theta, m^\Omega(\Gamma(A)) = m^\Theta(A),$$

with m superscript referring to the discernment frame on which the mass function is defined.

2.2.4. Combination rules

Different combination rules have been proposed to combine pieces of evidence defined on the same discernment frame. In particular, given two mass functions m_1 and m_2 representing two independent pieces of evidence, the result of Smets' conjunctive combination is defined for all $A \in 2^\Omega$ as:

$$(m_1 \odot m_2)(A) = \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega, \\ B \cap C = A}} m_1(B)m_2(C). \quad (2)$$

Smets' conjunctive combination is commutative, associative, has vacuous BBA as neutral element and the quantity $m_1 \odot m_2(\emptyset)$ is usually interpreted as the degree of conflict between m_1 and m_2 .

Among the alternative combination rules that provide a *normal* BBA, we consider Dubois-Prade's rule, defined for all $A \subseteq \Omega$ as:

$$(m_1 \star m_2)(A) = \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega, \\ B \cap C = A}} m_1(B)m_2(C) + \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega, \\ B \cap C = \emptyset, B \cup C = A}} m_1(B)m_2(C). \quad (3)$$

Dubois-Prade's rule combines the beliefs either conjunctively if they are consistent or disjunctively if there is conflict. It boils down to assuming that the conflict is due to the fact that one of the beliefs being false without knowing which one. Note that Dubois-Prade's rule is commutative, has the vacuous BBA as neutral element, but is not associative in the general case.

2.2.5. Decision making

After the evidence pieces have been represented and combined in 2^Ω , the final decision is usually made in Ω . The two most popular transformations used for retrieving a probabilistic measure from a BBA are the pignistic and the plausibility transformations expressed for all $\omega \in \Omega$ as:

$$BetP(\omega) = \frac{1}{1 - m(\emptyset)} \sum_{A \ni \omega} \frac{m(A)}{|A|}, \quad (4)$$

$$\pi(\omega) = \frac{pl(\omega)}{\sum_{\omega \in \Omega} pl(\omega)}, \quad (5)$$

where $|A|$ denotes the *cardinality* of A (i.e., the number of elements of Ω contained in A).

2.3. Basic Belief Assignment methods

Aside from methods where Basic Belief Assignments (BBAs), usually represented as mass functions, are set purely based on expert judgment, BBA construction methods in BFT are broadly categorized into similarity-based and optimization-based approaches. Similarity-based approaches construct BBAs by measuring the similarity between a new sample and predefined class representations

or models, utilizing various similarity metrics and data representations [32]. Optimization-based approaches formulate BBA problem as an objective criterion to be optimized, typically involving clustering consistency or classification accuracy. In the following subsections, we briefly review representative methods from each category, highlighting their distinct characteristics and contributions.

2.3.1. Similarity-based methods

Interval number-based methods construct intervals from statistical analysis or clustering, and infer BBAs by evaluating the similarity between samples and these intervals [50,71,39]. Core sample-based methods select representative samples, assigning belief based on proximity to these exemplars [69]. Triangular fuzzy number-based methods employ fuzzy membership functions, calculating fuzzy similarities or difference degrees to derive mass assignments, effectively addressing uncertainty and open-world scenarios [43,46,31]. Confusion matrix-based methods derive BBAs from the classifier's performance characteristics, using metrics such as accuracy and recall extracted from the confusion matrix to quantify and assign belief masses to each class [13]. K-means clustering combined with KNN methods construct BBAs through local similarities to clustered prototypes, enhancing robustness and efficiency [24]. Membership-function-based approaches construct class models with various parametric functions (e.g., Gaussian, triangular), to derive masses from membership degrees or classifier outputs. They often incorporate multiple similarity measures and entropy measures to refine belief assignments [30,32,70]. In an even more sophisticated way, [62] integrates Mahalanobis distance, cosine similarity, and belief entropy to refine BBAs and reduce conflict, explicitly accounting for both similarity to class prototypes and uncertainty in mass assignments.

2.3.2. Optimization-based methods

Evidential clustering methods, such as EVCLUS [20], ECM [44], and EK-NNclus [19], optimize criteria related to cluster conflict and consistency to generate credal partitions, encoding membership uncertainty [18]. Neural network extensions of evidential clustering (NN-EVCLUS) employ trained neural networks to directly learn mappings from feature spaces to mass functions, minimizing conflicts among similar samples and enhancing scalability and interpretability [17].

Other distance-based neural network layers further integrate learnable distance-to-mass transformations within neural architectures, explicitly optimizing mass assignments to achieve precise and robust classifications under uncertainty [14,16].

While significant research efforts have been devoted to mass function assignment, existing methods largely rely on data-driven similarity measures or optimization-based criteria, without explicitly incorporating structured domain knowledge. Consequently, these approaches do not exploit available expert insights and qualitative constraints that can significantly improve the reliability and interpretability of the assigned belief masses. In this paper, we address this gap by proposing a systematic framework that explicitly integrates structured domain knowledge through *Evidence Mapping Functions* (EMFs), subsequently refined by a *Constrained Neural Network* (Constrained-NN) optimization procedure guided by a task-specific loss function.

3. Proposed framework

The methodological contribution of this paper is to provide a systematic framework for defining reliable mass functions that both integrates domain knowledge and further tunes the mass representation through a learning process. The framework consists of three main stages (see Fig. 1):

- **Parametric representation of evidence:** Evidence Mapping Functions (EMFs) are defined based on expert knowledge to map modality-specific feature vectors to mass functions as a function of a set of hyperparameters to be optimized.
- **Constrained-NN representation of evidence:** The EMFs are then formulated as a Constrained-NN to enable the estimation of their hyperparameters. The prior constraints are formulated as mathematical relationships imposed on the network weights.
- **Loss function design:** A task-specific loss function is designed to optimize the hyperparameters of the EMFs.

In Fig. 1, the first stage is represented by the leftmost vertical block, which is informed by domain knowledge and outputs a constrained neural network. The second and third blocks correspond to the application of this network to data: the middle block receives feature vectors as input and outputs BBAs per modality, which are then fused in the final block to compute the task-specific loss.

Each stage is discussed in detail below, with an example provided to illustrate how the EMFs are represented as Constrained-NNs.

3.1. Parametric representation of evidence

To reliably represent evidence from information sources, we construct Evidence Mapping Functions that map feature information to mass functions based on some prior knowledge. Formally, an EMF Φ is defined as:

$$\Phi : \mathbb{R}^n \rightarrow [0, 1]^{2^{|\Omega|}}, \quad m = \Phi(\mathbf{x}),$$

where $\mathbf{x} \in \mathbb{R}^n$ is the feature vector of dimension n , m is the resulting mass function, and $2^{|\Omega|}$ denotes the number of elements in the powerset of Ω . These functions define the prior mass distributions in the feature space, based on domain-specific qualitative constraints, before any learning process: domain expertise and prior knowledge are encapsulated in their specific parametric expression.

Each EMF instance is defined by a set of hyperparameters, such as discounting coefficients, which will be optimized based on a given dataset. Since the optimization procedure is implemented within a NN framework, each EMF must be compatible with a neural representation—namely, expressible as a sequence of linear combinations followed by non-linear activation functions.

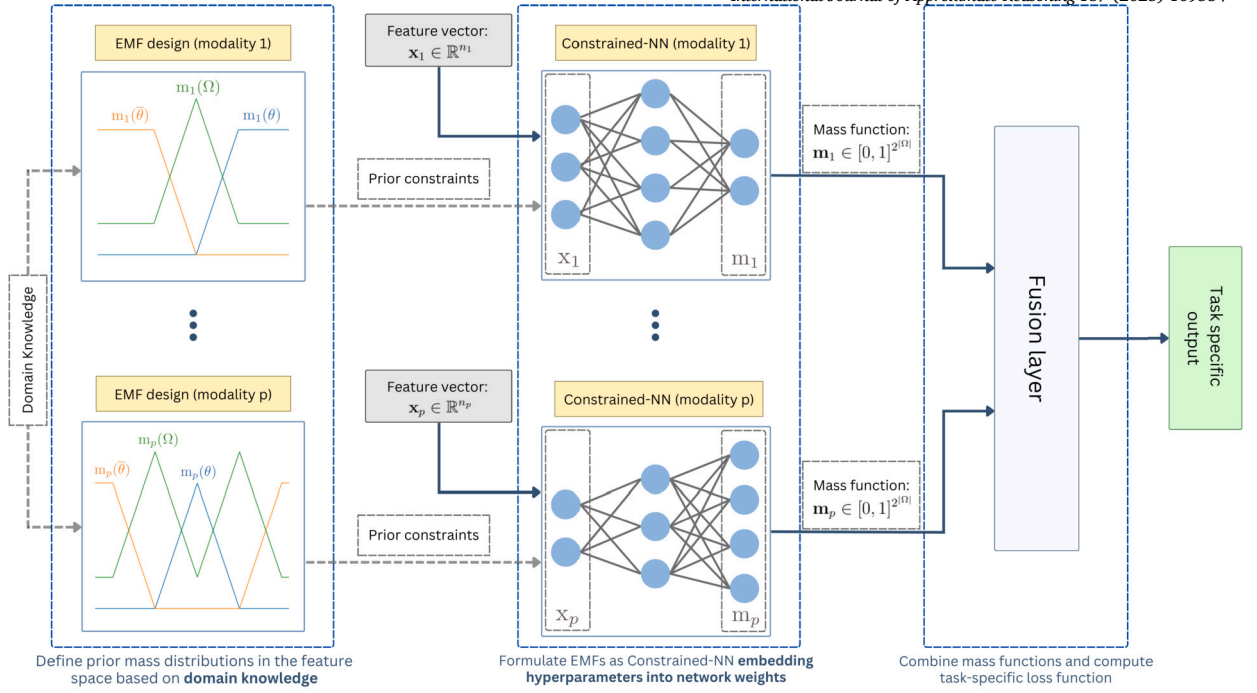


Fig. 1. The proposed multimodal framework for mass function assignment, where $i \in \llbracket 1, p \rrbracket$ is the modality index and $n_i, i \in \llbracket 1, p \rrbracket$ is the dimension of the feature vector of modality i .

It is worth noting that evidence, like other functions, can be decomposed using various parametric forms such as Gaussians, sigmoids, or tanh, each offering different shapes of mass distributions and enabling smooth transitions across the feature space. Alternatively, piecewise linear functions provide interpretability and greater flexibility, particularly when modeling distinct states across predefined feature intervals. These functions can be conveniently decomposed into a set of shifted ReLU functions—essentially ramp signals—making them naturally compatible with a Constrained-NN representation. Examples of such a piecewise linear function decomposition are provided in Toy Example 1 as well as in Section 4.

3.2. Constrained-NN representation of EMFs

Once the parametric expression of the EMFs has been set, they are expressed as Constrained-NNs while ensuring that their hyperparameters are embedded within the trainable weights of the network for optimization. Specifically, any EMF has to be expressed as:

$$\begin{cases} \Phi(\mathbf{x}) = f^{(l)}(\mathbf{W}^{(l)}\mathbf{a}_{l-1} + \mathbf{b}^{(l)}), \\ \mathbf{a}_j = f^{(j)}(\mathbf{W}^{(j)}\mathbf{a}_{j-1} + \mathbf{b}^{(j)}), \quad j \in \llbracket 1, l-1 \rrbracket \\ \mathbf{a}_0 = \mathbf{x}, \end{cases}$$

where \mathbf{x} denotes the feature vector, \mathbf{a}_{j-1} , $\mathbf{b}^{(j)}$ and $\mathbf{W}^{(j)}$ represent the output, weight matrix, and bias vector of layer j , respectively. The function $f^{(j)}$ is the activation function of layer j . The formulation of the EMFs as Constrained-NNs provides the basis for further refinement of mass assignments.

Toy example 1. Consider a temperature sensor used to infer the state of water, with two possible hypotheses: H_L , “The water is liquid”, and H_F , “The water is frozen”. The discernment frame is $\Omega = \{H_L, H_F\}$. Considering a normal BBA, i.e. $m(\emptyset) = 0$, the EMF assigns mass values to H_L , H_F and Ω based on the measured temperature τ , guided by the following qualitative constraints: confidence in H_L increases as τ rises above 0°C ; confidence in H_F increases as τ drops below 0°C ; uncertainty is highest when τ is close to 0°C . For the sake of brevity, we focus on the hypothesis H_L , whose mass value is defined as a piecewise linear function of τ :

$$m(H_L | \tau) = \begin{cases} 0 & \text{if } \tau \leq \tau_{\min}, \\ \frac{\alpha(\tau - \tau_{\min})}{\tau_{\max} - \tau_{\min}} & \text{if } \tau_{\min} < \tau \leq \tau_{\max}, \\ \alpha & \text{if } \tau > \tau_{\max}, \end{cases}$$

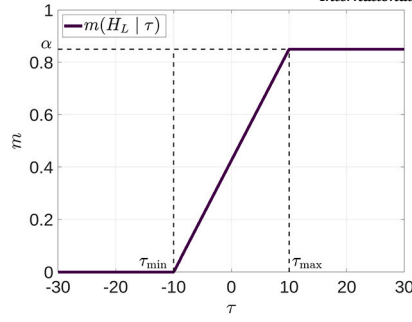


Fig. 2. The mass assigned to H_L as a function of temperature τ .

where the discounting coefficient α and the thresholds τ_{\min} and τ_{\max} need to be optimized. To implement this function in a NN, we use the ReLU function and its shifted version, which are defined as follows:

$$R(x) = \max(x, 0), \quad (6)$$

$$R_a(x) = \max(x - a, 0). \quad (7)$$

Specifically, $m(H_L | \tau)$ can be rewritten as the difference of two shifted ReLU outputs (see Fig. 2):

$$m(H_L | \tau) = \frac{\alpha}{\tau_{\max} - \tau_{\min}} \left(R_{\tau_{\min}}(\tau) - R_{\tau_{\max}}(\tau) \right).$$

Besides, the 2D vector $\mathbf{R} = \frac{\alpha}{\tau_{\max} - \tau_{\min}} \begin{bmatrix} R_{\tau_{\min}}(\tau) \\ R_{\tau_{\max}}(\tau) \end{bmatrix}$ can be written as:

$$\mathbf{R} = R \left(\begin{bmatrix} \frac{\alpha}{\tau_{\max} - \tau_{\min}} \\ \frac{\alpha}{\tau_{\max} - \tau_{\min}} \end{bmatrix} \cdot \tau + \begin{bmatrix} -\tau_{\min} \\ -\tau_{\max} \end{bmatrix} \frac{\alpha}{\tau_{\max} - \tau_{\min}} \right).$$

Setting $\alpha' = \frac{\alpha}{\tau_{\max} - \tau_{\min}}$, $b_1 = -\tau_{\min}\alpha'$ and $b_2 = -\tau_{\max}\alpha'$, we obtain the Constrained-NN representation for H_L as a two-layers network:

$$\begin{cases} \mathbf{R} = R \left(\begin{bmatrix} \alpha' \\ \alpha' \end{bmatrix} \cdot \tau + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \\ m(H_L | \tau) = \begin{bmatrix} 1 & -1 \end{bmatrix} \mathbf{R} \end{cases}$$

The learnable parameters are the shared weight α' and the biases b_1 and b_2 . The constraints are the facts that α' is shared and positive, and that the biases are negative.

3.3. Loss function design

In this third stage, the mass functions generated by the EMFs in their Constrained-NN form are combined in a final layer, from which the loss function is derived. Since loss functions are inherently task-dependent, this section primarily provides and discusses guidelines on the use of BFT operators or measures to assist users in defining loss functions tailored to their application. These guidelines are illustrated with a comprehensive example in the next section. Key measures that can be used to define the loss function include:

3.3.1. Evidence inconsistency

Inconsistency measures are particularly interesting for multimodal fusion tasks, as they quantify the overall coherence of the combined mass function. In BFT, the internal inconsistency of a mass function can be quantified using various approaches. Two broad categories include extensions of the Shannon entropy formula to belief functions, and measures that evaluate the consistency between sets in 2^Ω . Table 1 presents a selection of (in)consistency measures based on both approaches.

3.3.2. Non-specificity and total uncertainty

For tasks requiring more committed mass functions, uncertainty metrics such as non-specificity can be minimized. Non-specificity measures focus on the cardinality of the focal elements, assigning a maximal value (least specific) to the vacuous mass function ($m_v(\Omega) = 1$) and a minimal value (most specific) to a Bayesian mass function (i.e., all focal elements have a cardinality of 1). These measures are particularly useful for capturing the degree of imprecision in a mass function. A selection of non-specificity measures is presented in Table 2.

To jointly capture the two dimensions of uncertainty—namely, inconsistency and non-specificity—a simple way is to perform a weighted combination of the corresponding measures:

Table 1

Some measures of (in)consistency based on entropy measures and consistency between sets.

Entropy measures		Consistency between sets	
Author(s)	Definition	Author(s)	Definition
Höhle [27]	$-\sum_{A \subseteq \Omega} m(A) \log Bel(A)$	Jousselme et al. [34]	$\sum_{A \subseteq \Omega} m(A) \sum_{B \subseteq \Omega} m(B) \frac{ A \cap B }{ A \cup B }$
Yager [66]	$-\sum_{A \subseteq \Omega} m(A) \log Pl(A)$	Destercke & Burger [21]	$\max_{\omega \in \Omega} Pl(\{\omega\})$
Nguyen [45]	$-\sum_{A \subseteq \Omega} m(A) \log m(A)$	Destercke & Burger [21]	$1 - m(\emptyset) = \max_{A \subseteq \Omega} Pl(A)$

Table 2

Various metrics for evaluating the non-specificity of a mass function.

Authors	Name	Definition
Dubois & Prade [22]	Non-specificity	$\sum_{A \subseteq \Omega} m(A) \log A $
Lamata & Moral [37]	Imprecision	$\log \left(\sum_{A \subseteq \Omega} m(A) A \right)$
Dubois & Prade [23]	Generalized cardinality	$\sum_{A \subseteq \Omega} m(A) A $

$$v(m) = av_c(m) + bv_s(m), \quad (8)$$

where $(a, b) \in \mathbb{R}^2$, and v_c and v_s denote an inconsistency measure and a non-specificity measure, respectively, chosen from those listed in Tables 1 and 2. This total uncertainty formulation can be incorporated into the loss function to ensure coherent fusion while maintaining a desired level of commitment.

3.3.3. Decision-making models

In BFT, decisions are made under partial knowledge, typically modeled through an uncertainty function and a utility function that quantifies the desirability of outcomes given the possible states of the world (i.e., the elements of the frame of discernment). Further details on decision-making within the BFT framework can be found in [15]. In supervised classification tasks, mass-to-probability transformations—such as those described in Eq. (4) and (5)—can be applied to convert belief assignments into probability distributions. Once these probability distributions are obtained, standard probabilistic loss functions, such as categorical cross-entropy or Kullback–Leibler divergence, can be used for training.

4. Application on pigment identification

This section provides a real-world application of the framework proposed in Section 3, with the application context outlined in Section 4.1. The remainder of the section focuses on the implementation of each stage of the framework.

4.1. Application context

The fusion module proposed in this section is designed to identify the pigments used in photographs based on the *trichromatic principle*.² Specifically, it integrates information from XRF maps and optical imaging modalities, namely VIS images, IRFC images and UVFC images, all represented in the *CIELAB* color space. The multimodal approach leverages the distinctive combination of pigment-discriminant features across different modalities to achieve a more robust identification. Note, however, that since the photographs are created by superposing layers of different colors, each layer is considered independently. As a result, the pigment identification problem is addressed separately for each layer, i.e., for each color.

If pigment identification is used for artwork dating, both False Positives (FP) and False Negatives (FN) can mislead historical interpretation or result in an inaccurate reconstruction of the artist's palette. For example, in the analysis of early color photographs, distinguishing between Prussian blue and other blue pigments is crucial for understanding the technical processes involved in his photomechanical experiments.

4.2. EMF design

Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \omega_\emptyset\}$ be the discernment frame of our application for a given pigment color, where $\omega_i, i \in \llbracket 1, n \rrbracket$, represent the n pigments associated with the considered color and ω_\emptyset denotes the hypothesis of the absence of all considered pigments. Let also $\Omega^* = \Omega \setminus \{\omega_\emptyset\}$ be the disjunction of all pigments considered in Ω . Multiple pigments may present undistinguishable features in

² The trichromatic principle is a color theory concept where a wide range of colors is obtained by mixing three primary colors—red, blue, and yellow, either by direct blending or superposition.

a given modality. Consequently, we define a coarsening of Ω for each modality such that $\Theta_{mod} = \{\theta_{mod,1}, \theta_{mod,2}, \dots, \theta_{mod,k}, \omega_\emptyset\}$, where $\theta_j, \forall j \in \llbracket 1, k \rrbracket$, represent the k groups of (one or more) pigments that share similar features in a modality mod .

Our objective is to derive a segmentation of a given color layer in terms of pigments. To enhance robustness against acquisition noise and spatial misalignment between modalities, this segmentation is performed at a superpixel level, with mass functions assigned to each superpixel. Furthermore, information derived from XRF maps and optical imaging modalities is inherently heterogeneous, as each describes distinct physical properties that require specific interpretation rules. Consequently, distinct EMFs are constructed for each type of modality, mapping modality-specific features to mass functions. In general terms, the definition of an EMF for a given source of information boils down to answering the following questions:

1. What features are used to characterize each source?
2. What are the expected focal elements?
3. How does the mass distribution over the focal elements vary as a function of the considered features?
4. What are the EMF learnable hyperparameters?

Fig. 4 synthesizes the process of using expert knowledge to construct modality-specific EMFs by answering these questions.

For an XRF map describing the distribution of a given element el in a photograph (cf. Section 4.1) ph :

- The feature vector consists of the elemental concentration: $\mathbf{x} = c_{el}$.
- The focal elements of the resulting mass function are the set of pigments θ_{el} associated with element el , its complement $\overline{\theta_{el}}$, as well as the Ω .
- The mass assigned to θ_{el} increases proportional to c_{el} while the mass assigned to $\overline{\theta_{el}}$ decreases proportional to c_{el} and $m(\Omega)$ is set using the sum-to-one constraint of mass functions (see Fig. 3a).
- Domain knowledge suggests a strong negative correlation between pigment concentration and artwork lightness l_{vis} : lighter regions reflect the artwork's background, while darker regions correspond to areas with high pigment concentrations. Based on this principle, the source reliability α_{el} is defined as proportional to the absolute value of the correlation r_{el}^{ph} between the elemental concentration and the lightness component (in the LAB color space) of the VIS modality of the same photograph: $\alpha_{el}(r_{el}^{ph}) = \beta_{el} |r_{el}^{ph}|$, with β_{el} a learnable parameter. Then if this correlation exceeds a predefined threshold (determined experimentally), the mass distribution will take the form illustrated in Fig. 3b.
- The (super)pixel intensity of XRF maps does not provide absolute elemental concentration values but rather indicates relative concentrations across an artwork. Based on domain knowledge, we assume that image lightness is a more consistently calibrated quantity across artworks. Consequently, the hyperparameters $\overline{\rho_{el}}, \rho_{el}$ are inferred from the established relationship between elemental concentration and lightness of the VIS image, which is achieved by projecting the learnable parameters $\underline{\lambda_{el}}, \overline{\lambda_{el}}$ onto c_{el} (see Fig. 3c).

For optical modalities:

- Super-pixel features consist of the lightness and hue³ values: $\mathbf{x} = [l_{opt} \quad h_{opt}]^T$.
- Optical modalities provide richer information than XRF, as they capture data on all considered pigment groups. In this work, each piece of information related to a given pigment group $\theta_{opt} \in \Theta_{opt}$ is represented by a θ_{opt} mass function. These θ_{opt} mass functions are then combined using the Dubois-Prade combination rule to derive the opt mass function, gathering all evidences from the modality observation.
- For each θ_{opt} mass, the set of focal elements is: $F_{m_{opt}} = \{\theta_{opt}, \overline{\theta_{opt}}, \Omega^*, \Omega\}$.
- According to domain knowledge, darker regions support the presence of pigments without the ability to distinguish them. As a result, an important mass value is assigned to Ω^* in darker regions, while lighter areas are characterized by a high mass assigned to $\overline{\theta_{opt}}$ (see Fig. 3e and 3f).
- A high mass is assigned to θ_{opt} when h_{opt} takes values close to the characteristic hue of the group of pigments θ_{mod} (Fig. 3d).
- $m_{opt}(\Omega)$ is obtained using the sum-to-one constraint (see Fig. 3g).

4.3. Constrained-NN representation of EMFs

The Constrained-NN representation of the EMF of each type of modality is derived by expressing the piecewise linear functions in Fig. 3 as a series of linear combinations followed by ReLU activations (using Eq. (6) and (7)).

³ $h_{opt} = \arctan\left(\frac{a_{opt}}{b_{opt}}\right)$, with a_{opt}, b_{opt} the A and B channels in the LAB color space.

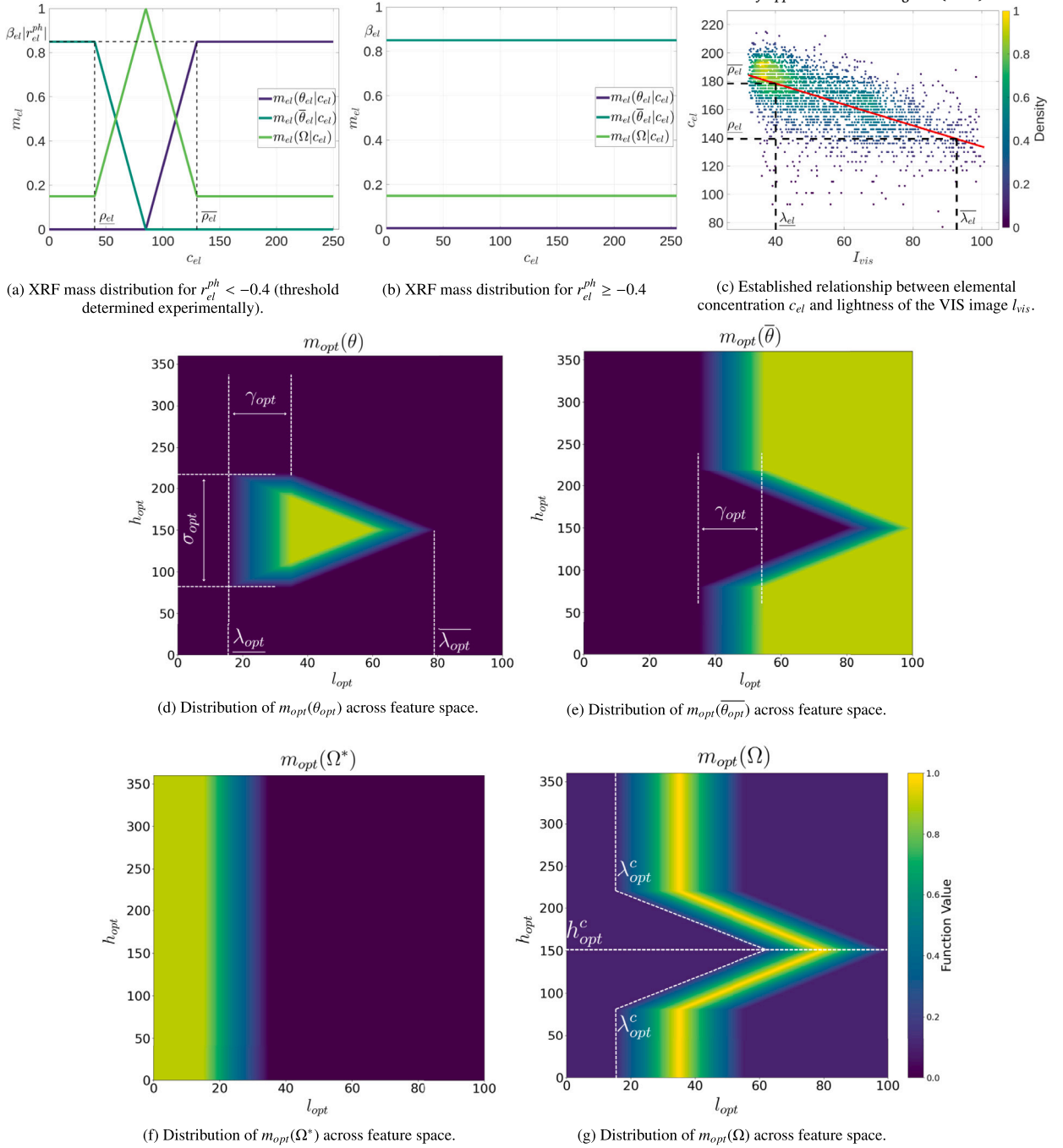


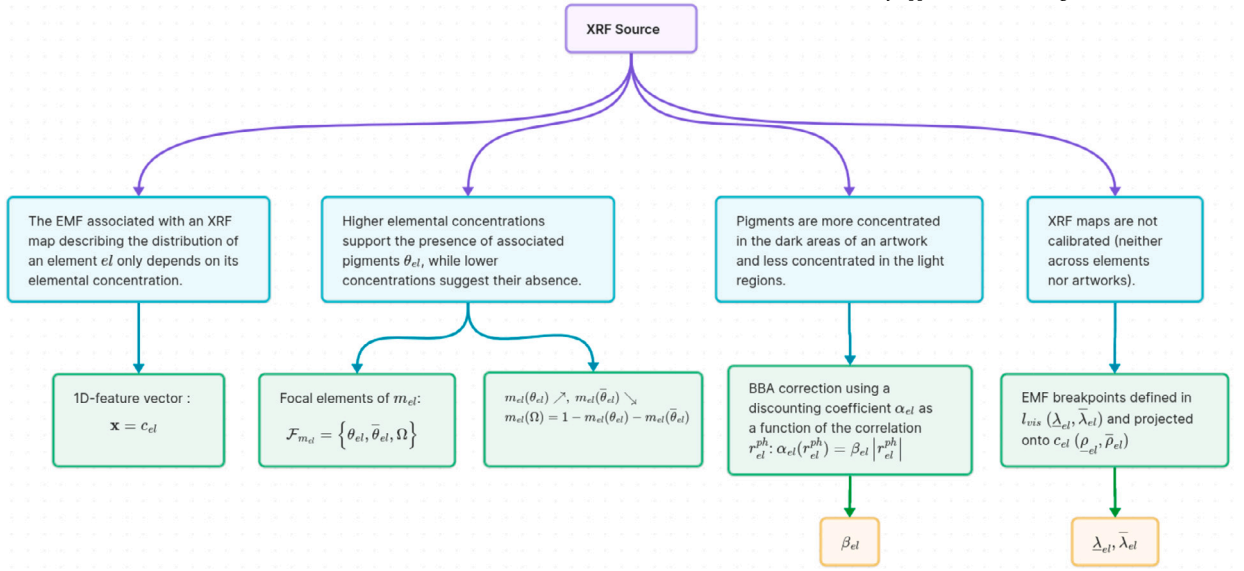
Fig. 3. Mass distributions for XRF and optical imaging modalities across their respective feature spaces; the specific learnable hyperparameters used are drawn on each figure. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4.3.1. XRF modalities

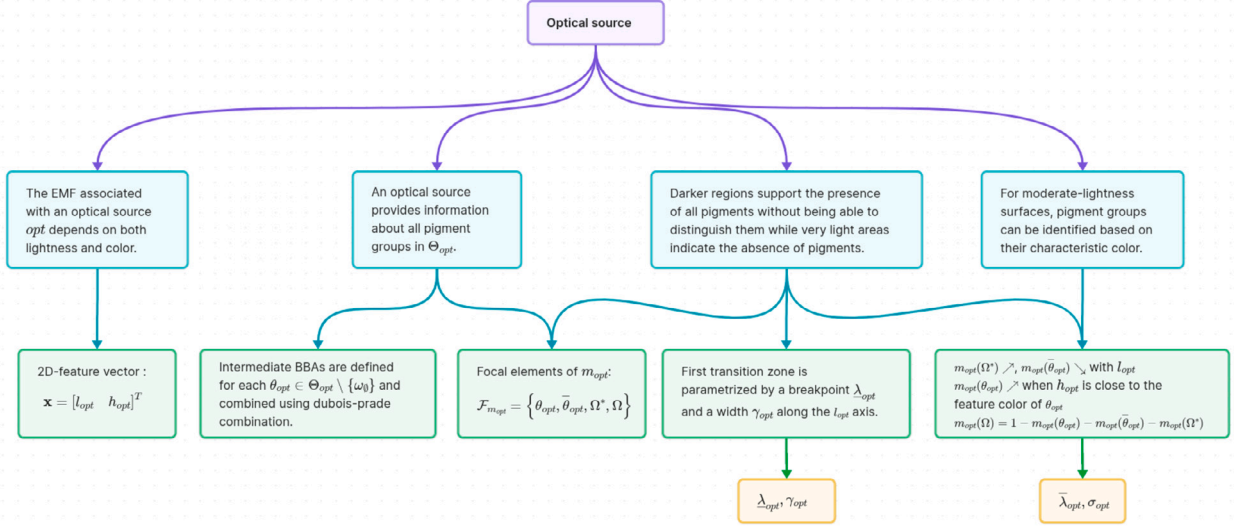
From observed concentration c_{el} , we derive the 3D vector:

$$\begin{bmatrix} s_1(c_{el}) \\ s_2(c_{el}) \\ s_3(c_{el}) \end{bmatrix} = \frac{2\beta_{el}|r_{el}^{ph}|}{\rho_{el} - \bar{\rho}_{el}} \begin{bmatrix} R_{\rho_{el}^m}(c_{el}) \\ R_{\bar{\rho}_{el}}(c_{el}) \\ R_{\rho_{el}}(c_{el}) \end{bmatrix},$$

where $\rho_{el}^m = \frac{1}{2}(\rho_{el} + \bar{\rho}_{el})$. Then, from Fig. 3a:



(a) EMF design for XRF modalities



(b) EMF design for optical imaging modalities

Fig. 4. Hierarchical tree representation of the EMF design for each modality. The purple node indicates the modality type, blue nodes specify the qualitative constraints, green nodes illustrate how expert guidelines influence EMF design choices, and yellow nodes represent the resulting learnable hyperparameters.

$$m_{el}(\theta_{el}) = s_1(c_{el}) - s_2(c_{el}),$$

$$m_{el}(\overline{\theta_{el}}) = R\left(\beta_{el}|r_{el}^{ph}| - s_3(c_{el})\right),$$

$$m_{el}(\Omega) = 1 - m_{el}(\theta_{el}) - m_{el}(\overline{\theta_{el}}).$$

As explained in Section 4.2, the thresholds $\underline{\rho_{el}}$ and $\overline{\rho_{el}}$ are obtained by projecting the learnable parameters $\overline{\lambda_{el}}$ and $\underline{\lambda_{el}}$ using the regression line illustrated in Fig. 3c, so that:

$$\underline{\rho_{el}} = a_{el}^{ph} \cdot \overline{\lambda_{el}} + b_{el}^{ph}, \quad \overline{\rho_{el}} = a_{el}^{ph} \cdot \underline{\lambda_{el}} + b_{el}^{ph},$$

with a_{el}^{ph} and b_{el}^{ph} the slope and offset of the regression line, respectively. Note that both of these parameters are photograph (or artwork) and element specific.

In order to ensure the embedding of the EMF's hyperparameters within the trainable weights of the network, we consider the following variable changes:

$$\varphi_{el} = \frac{\beta_{el}}{\lambda_{el} - \lambda_{el}}, \quad \rho'_{el} = \frac{\rho_{el}|r_{el}^{ph}|}{a_{el}^{ph}}, \quad \overline{\rho'}_{el} = \frac{\overline{\rho_{el}}|r_{el}^{ph}|}{a_{el}^{ph}},$$

so that we obtain the Constrained-NN representation:

$$\begin{aligned} \begin{bmatrix} \rho'_{el} \\ \overline{\rho'}_{el} \end{bmatrix} &= \begin{bmatrix} \lambda_{el} \\ \overline{\lambda_{el}} \end{bmatrix} \cdot |r_{el}^{ph}| + \begin{bmatrix} \frac{b_{el}^{ph}|r_{el}^{ph}|}{a_{el}^{ph}} \\ \frac{b_{el}^{ph}|r_{el}^{ph}|}{a_{el}^{ph}} \end{bmatrix}, \\ \begin{bmatrix} s_1(c_{el}) \\ s_2(c_{el}) \\ s_3(c_{el}) \end{bmatrix} &= R \begin{bmatrix} -\varphi_{el} & -\varphi_{el} & \varphi_{el} \\ 0 & -2\varphi_{el} & \varphi_{el} \\ -2\varphi_{el} & 0 & \varphi_{el} \end{bmatrix} \begin{bmatrix} \rho'_{el} \\ \overline{\rho'}_{el} \\ \frac{2c_{el}|r_{el}^{ph}|}{a_{el}^{ph}} \end{bmatrix}, \\ \begin{bmatrix} m_{el}(\theta_{el}) \\ m_{el}(\overline{\theta_{el}}) \end{bmatrix} &= R \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} s_1(c_{el}) \\ s_2(c_{el}) \\ s_3(c_{el}) \end{bmatrix} + \begin{bmatrix} 0 \\ \beta_{el}|r_{el}^{ph}| \end{bmatrix}, \\ \begin{bmatrix} m_{el}(\theta_{el}) \\ m_{el}(\overline{\theta_{el}}) \\ m_{el}(\Omega) \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} m_{el}(\theta_{el}) \\ m_{el}(\overline{\theta_{el}}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

4.3.2. Optical imaging modalities

As explained in Section 4.2, the EMF of an optical imaging modality is constructed by first defining an intermediary mass function for each $\theta_{opt} \in \Theta_{opt}$, and combining them using the Dubois-Prade rule. Since the order of combination of the sources is irrelevant, associativity is required. **Proposition 1** in Appendix B establishes the associativity of the Dubois-Prade rule under conditions that are met in our application. Besides, they are general enough to apply to other setups. Note also, that the Constrained-NN representation of optical modality belief provides an example with BBAs depending on 2D feature vectors that can also find applications in other domains that cultural heritage.

For a given $\theta_{opt} \in \Theta_{opt}$, the distribution of mass values of the corresponding BBA across the feature space $\mathbf{x} = [l_{opt} \ h_{opt}]^T$ is illustrated in Fig. 3d-3g. From Fig. 3g, we can write the lightness breakpoint λ_{opt}^c as:

$$\lambda_{opt}^c(h_{opt}) = l_1(h_{opt}) - 2l_2(h_{opt}) + l_3(h_{opt}) + \lambda_{opt},$$

with:

$$\begin{bmatrix} l_1(h_{opt}) \\ l_2(h_{opt}) \\ l_3(h_{opt}) \end{bmatrix} = \frac{\overline{\lambda_{opt}} - \lambda_{opt}}{\sigma_{opt}} \begin{bmatrix} R_{h_{opt}^c - \sigma_{opt}}(h_{opt}) \\ R_{h_{opt}^c}(h_{opt}) \\ R_{h_{opt}^c + \sigma_{opt}}(h_{opt}) \end{bmatrix}.$$

Then, setting:

$$\begin{bmatrix} s_1(l_{opt}) \\ s_2(l_{opt}) \\ s_3(l_{opt}) \\ s_4(l_{opt}) \end{bmatrix} = \frac{\alpha_{opt}}{\gamma_{opt}} \begin{bmatrix} R_{\lambda_{opt}}(l_{opt}) \\ R_{\lambda_{opt}^c}(l_{opt}) \\ R_{\lambda_{opt}^c + \gamma_{opt}}(l_{opt}) \\ R_{\lambda_{opt}^c + 2\gamma_{opt}}(l_{opt}) \end{bmatrix},$$

the mass values represented in Fig. 3d-3g can be derived:

$$\begin{aligned} m_{opt}(\Omega^*) &= R(\alpha_{opt} - s_1(l_{opt})) \\ m_{opt}(\Omega) &= s_2(l_{opt}) - 2s_3(l_{opt}) + s_4(l_{opt}) + (1 - \alpha_{opt}) \\ m_{opt}(\overline{\theta_{opt}}) &= s_3(l_{opt}) - s_4(l_{opt}) \\ m_{opt}(\theta_{opt}) &= 1 - m_{opt}(\Omega) - m_{opt}(\overline{\theta_{opt}}) - m_{opt}(\Omega^*). \end{aligned}$$

Similarly to XRF-EMFs, the following variable changes ensure the embedding of the EMF's hyperparameters within the trainable weights of the network:

$$\begin{aligned} \Delta\lambda_{opt} &= \overline{\lambda_{opt}} - \lambda_{opt}, \quad \delta_{opt} = \frac{\Delta\lambda_{opt}}{\sigma_{opt}}, \quad \varphi_{opt} = \frac{\alpha_{opt}}{\gamma_{opt}}, \\ \psi_{opt} &= \varphi_{opt} \lambda_{opt}. \end{aligned}$$

Consequently, the Constrained-NN representation of the EMF related to one pigment group $\theta_{opt} \in \Theta_{opt}$ in a given optical imaging modality is given by (for conciseness, we omit dependent variables: for example, $l_1(h_{opt})$ will be written as l_1):

$$\begin{aligned} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} &= R \begin{bmatrix} \delta_{opt} & -\delta_{opt} \\ \delta_{opt} & -\delta_{opt} \\ \delta_{opt} & -\delta_{opt} \end{bmatrix} \begin{bmatrix} h_{opt} \\ h_{opt}^c \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \Delta \lambda_{opt} \\ \lambda_{opt}^c &= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} + \lambda_{opt} \\ \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix} &= R \begin{bmatrix} 0 & \varphi_{opt} \\ -\varphi_{opt} & \varphi_{opt} \\ -\varphi_{opt} & \varphi_{opt} \\ -\varphi_{opt} & \varphi_{opt} \end{bmatrix} \begin{bmatrix} \lambda_{opt}^c \\ l_{opt} \end{bmatrix} + \begin{bmatrix} -\psi_{opt} \\ 0 \\ -\alpha_{opt} \\ -2\alpha_{opt} \end{bmatrix} \\ \begin{bmatrix} m_{opt}(\overline{\theta_{opt}}) \\ m_{opt}(\Omega^*) \\ m_{opt}(\Omega) \end{bmatrix} &= R \begin{bmatrix} 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha_{opt} \\ 1 - \alpha_{opt} \end{bmatrix} \\ \begin{bmatrix} m_{opt}(\overline{\theta_{opt}}) \\ m_{opt}(\overline{\theta_{opt}}) \\ m_{opt}(\Omega^*) \\ m_{opt}(\Omega) \end{bmatrix} &= \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{opt}(\overline{\theta_{opt}}) \\ m_{opt}(\Omega^*) \\ m_{opt}(\Omega) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

4.4. Loss function design

The mass functions obtained from each modality in stage 2 (see Section 4.3) are fused in a final layer using the conjunctive combination rule defined in Eq. (2). In our application, an effective representation of evidence is expected to exhibit both consistency across sources and accuracy in classification. To encourage these properties, several loss functions were investigated under both supervised and unsupervised settings.

Let m_c^i denote the combined mass function of the i^{th} training example. In the unsupervised setting, we introduce a loss function that minimizes the total uncertainty of m_c^i , combining an inconsistency term and a non-specificity term (see Tables 1 and 2 for common formulations of these measures) as defined in Eq. (8), resulting in the following loss function:

$$\mathcal{L}_{\text{unsup}} = \frac{1}{N} \sum_{i=1}^N \left(m_c^i(\emptyset) + \sum_{A \subseteq \Omega} m_c^i(A) |A| \right),$$

where N is the number of training examples. This unsupervised loss encourages both low conflict and high specificity in the fused mass function but lacks the supervision required to ensure accurate class prediction. To address this, we introduce a supervised extension by adding a classification loss term based on the pignistic transformation of m_c (see Eq. (4)). The resulting loss function is defined as:

$$\mathcal{L}_{\text{sup-1}} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{|\Omega|} y_{ij} \log(\hat{y}_{ij}) + \lambda \cdot \mathcal{L}_{\text{unsup}}, \quad (9)$$

where λ is a weighting coefficient, empirically set to 0.1, y_{ij} denotes the true belonging coefficient for class j (i.e., component j of one-hot vector) of sample i , and \hat{y}_{ij} is the predicted probability for that class obtained from the pignistic transformation of m_c^i , computed as:

$$\hat{y}_{ij} = \frac{1}{1 - m_c^i(\emptyset)} \sum_{\substack{A \subseteq \Omega \\ A \ni \omega_j}} \frac{m_c^i(A)}{|A|}.$$

A more compact loss function was also evaluated, aiming to simultaneously ensure fusion coherence and classification accuracy. This loss leverages a modified version of the pignistic transformation (Eq. (4)) in which the normalization factor is omitted:

$$\mathcal{L}_{\text{sup-2}} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{|\Omega|} y_{ij} \log \left(\sum_{\substack{A \subseteq \Omega \\ A \ni \omega_j}} \frac{m_c^i(A)}{|A|} \right). \quad (10)$$

By removing the normalization term $1 - m_c^i(\emptyset)$, the resulting output vector no longer sums to one. Since the target vector \mathbf{y}_i remains a valid probability distribution (i.e., sums to one), the optimization implicitly encourages the predicted vector to do the same.

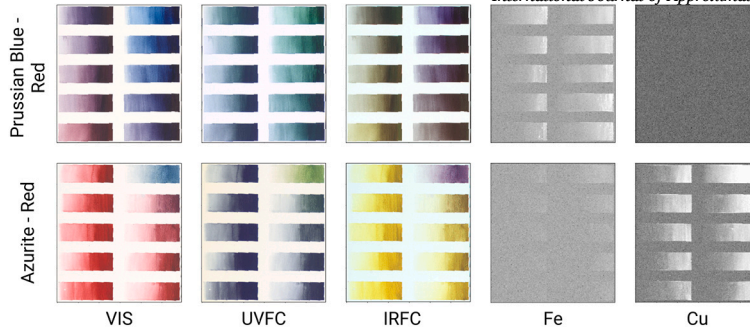


Fig. 5. Examples of mock-up samples containing Prussian Blue and Azurite mixed with a red pigment. The first row showcases the output of all considered modalities for mock-up samples containing Prussian blue mixed with a red pigment, whereas the second row depicts the same modalities for a mixture of Azurite and the same red pigment.

Consequently, this penalizes high conflict values $m_c^i(\emptyset)$ in the fused mass function—promoting consistency between sources—while still preserving the classification objective.

5. Experimental results

In this section, the proposed framework is experimentally validated. The experimental setups are first described, including the datasets, pigment classes, and training configuration. The framework is then evaluated on synthetic data and on mock-up samples. The synthetic data analysis includes a hyperparameter convergence study and a sensitivity analysis to both noise and dataset size. Finally, the loss functions presented in Section 4.4 are compared, and the classification results are analyzed at class level through ablation studies assessing dataset complexity, multimodality, and the role of optimization, followed by a comparison to state-of-the-art (SOTA) evidential classifiers.

5.1. Experimental setups

5.1.1. Considered pigments

To evaluate the proposed framework, we focus on blue pigment classification using mock-up samples representative of typical mixtures found in 19th-century photographic processes [29]. Based on expert knowledge, the considered blue pigments are Prussian Blue, Azurite, and Indigo. The discernment frame is defined as $\Omega = \{\omega_{pb}, \omega_{az}, \omega_{ind}, \omega_{\emptyset}\}$, where ω_{pb} , ω_{az} , and ω_{ind} correspond to the target pigments, and ω_{\emptyset} denotes the absence of blue pigments.

5.1.2. Datasets

Given the scarcity of labeled data in CH, we train and evaluate our model in two distinct settings: one using synthetic data and another using mock-up samples, where each experimental setup serves a specific objective. Indeed, first, training on synthetic data illustrates the convergence of the estimated EMF hyperparameters to their target values, proving the validity of the approach. Then, training on mock-up samples provides a more realistic assessment of the model classification performance, as mock-ups better simulate real-world conditions, effectively demonstrating the applicability of the approach to real-world scenarios.

Since we do not have a direct model to generate features from a given class, we take the opposite approach: we randomly sample feature values within each modality's feature space—lightness and hue for optical images, and elemental concentration for XRF maps. However, this random sampling can produce unrealistic combinations. To filter out such inconsistencies, we compute the ground truth class using the forward model with a predefined set of target hyperparameters, and discard any sample that results in a combined mass function with non-zero conflict. This ensures coherence between the sources of information. The primary objective of the training is to assess whether the estimated hyperparameters successfully converge to their target values.

The mock-ups dataset consists of mock-up paint samples of all possible mixtures of two pigments of different colors (blue, red, yellow). The considered pigments include three blue pigments (Prussian Blue, Azurite, and Indigo), three yellow pigments, and one red pigment, resulting in 15 different mixtures. Fig. 5 shows the optical images and elemental maps for Fe and Cu of the mock-up samples, corresponding to two mixtures containing Prussian Blue and Azurite, respectively. The pigment proportions in all the mixtures present in the dataset vary systematically from 100%–0% or 90%–10% to 10%–90%, decreasing and increasing in increments of 10%.

5.1.3. Considered class ambiguities

We consider two cases for handling class ambiguities: one where class ambiguities are explicitly modeled using coarse discernment frames (synthetic dataset) and one where they are only implicitly processed (mock-up dataset).

In the synthetic dataset, the following coarse discernment frames were defined for optical imaging modalities: $\Theta_{vis} = \{\{\omega_{pb}, \omega_{az}, \omega_{ind}\}, \omega_{\emptyset}\}$, $\Theta_{uvfc} = \{\{\omega_{pb}, \omega_{az}\}, \omega_{ind}, \omega_{\emptyset}\}$, and $\Theta_{irfc} = \{\omega_{pb}, \{\omega_{az}, \omega_{ind}\}, \omega_{\emptyset}\}$. These discernment frames reflect the assumption that some pigments should be indistinguishable under specific imaging modalities: blue pigments are indistinguishable

Table 3

Characteristic features of blue pigments across modalities.

Modality	Prussian Blue	Azurite	Indigo
h_{vis}^c (°)	286	243	300
h_{uvfc}^c (°)	172	120	150
h_{irfc}^c (°)	337	330	32
Chemical element	Iron (Fe)	Copper (Cu)	N/A (organic)

Table 4Summary of the performance metrics of the model (mean \pm standard deviation) trained on the synthetic dataset for different loss functions.

	CCE	MAE	MSE	KL-div
Δ_{err} (%)	0.43 \pm 0.41	1.58 \pm 4.73	2.7 \pm 5.28	11.9 \pm 3.91
Acc (%)	99.86 \pm 0.208	99.73 \pm 0.321	99.42 \pm 0.489	99.22 \pm 0.321

in VIS images, and the theoretical characteristic colors suggest that ω_{pb} and ω_{az} cannot be differentiated in UVFC images, while ω_{az} and ω_{ind} are indistinguishable in IRFC images.

However, upon analyzing the pigments used in the mock-ups, we observed subtle but noticeable differences in their characteristic feature colors. This suggests that, in practice, these pigments are more distinguishable than initially expected. As a result, when training on the mock-up dataset, it proved more practical and effective to avoid coarsening. Nonetheless, in the case of synthetic data, maintaining coarsening remains valuable, as it allows us to explore scenarios with significant class ambiguities, which can be useful for assessing the robustness of the model.

The characteristic features of the blue pigments used in the mock-up samples are reported in Table 3, where h_{opt}^c denotes the characteristic hue angle (in degrees) for each optical imaging modality.

5.1.4. EMF instantiation

Based on the pigment characteristics outlined in Table 3, we consider five sources: VIS, UVFC, and IRFC images, along with Fe and Cu maps, and an EMF is instantiated for each source.

5.1.5. Pre-processing

The XRF maps and optical images are acquired separately using different sensors, resulting in images with different resolutions and orientations. To address this problem, a perspective transformation is applied to XRF maps with nearest-neighbor interpolation to match the size and orientation of optical images. Additionally, the high noise present in XRF maps is addressed by applying a superpixel segmentation using the SLIC algorithm [1] resulting in homogeneous small regions in terms of color and texture, and each small region is represented by its median value in all modalities.

5.1.6. Training configuration

The datasets were split into training (80%), validation (10%), and test (10%) subsets. The model was trained using the Adam optimizer for 300 epochs with early stopping (patience of 15 epochs). The learning rates were empirically tuned and set to 5×10^{-5} for the mock-up dataset and 7×10^{-4} for the synthetic dataset with a learning rate decay factor of $e^{-0.22}$ applied every 40 epochs in both cases. The batch size was set to 32 for the mockup dataset. As for the synthetic dataset, the batch size varies with the size of the dataset (see the last paragraph of Section 5.2.2). Finally, the weight initialization strategy was consistent across both datasets: random initialization centered around an expert-informed “good” prior estimate, to facilitate faster and more stable convergence.

5.2. Results on the synthetic dataset

5.2.1. Hyperparameter convergence analysis

Table 4 summarizes the performance results obtained from training on the synthetic dataset using different loss functions (Categorical Cross-Entropy (CCE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Kullback-Leibler Divergence (KL-div)). For each loss function, the convergence of the model was evaluated on five sets of target hyperparameters, with five random initializations per set. To evaluate performance, we use two key metrics:

- The average relative estimation error of the hyperparameters across the EMFs of all sources. We recall definition of the relative estimation error as:

$$\Delta_{err} = \frac{|h - \hat{h}|}{h} \times 100,$$

where h and \hat{h} denote the target and estimated hyperparameters, respectively. The relative estimation errors are then averaged over the different hyperparameters and different runs.

- The classification accuracy Acc of the model averaged over the different classes and different runs.

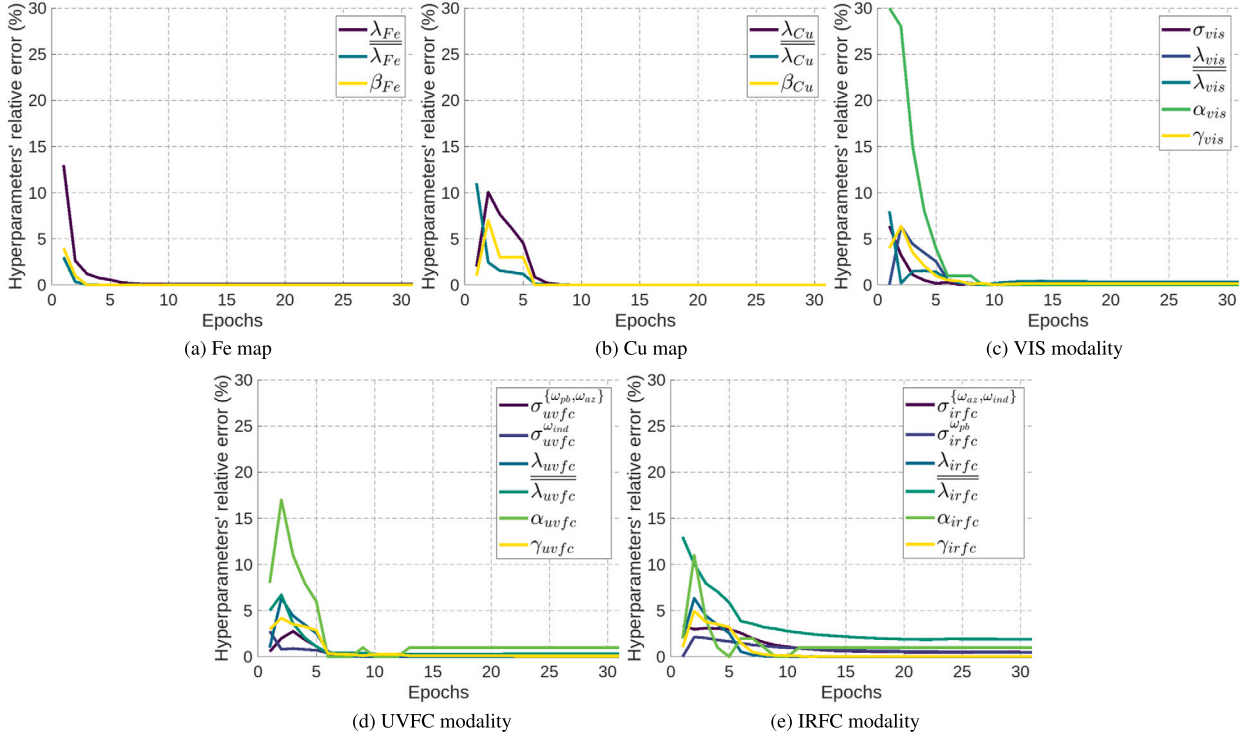


Fig. 6. Evolution of the relative estimation error for the hyperparameters of the EMFs across all considered modalities.

Table 5

Classification accuracy (*mean ± std*) for different levels of added Gaussian noise. The noise standard deviation is expressed as a percentage of the dynamic range of each input feature. Results are averaged across dataset sizes $\in \{10K, 100K, 1M, 10M\}$.

Noise std (%)	0	2	5	10
Accuracy (%)	100.00 ± 0.00	96.03 ± 0.07	87.40 ± 0.05	69.87 ± 0.34

The results indicate that the model is capable of estimating target hyperparameters with high precision, achieving an average relative estimation error as low as 0.43% with the CCE loss, while maintaining an accuracy of 99.86%. Moreover, CCE consistently outperforms the other loss functions, achieving the best results in both classification accuracy and hyperparameter estimation precision. Even for relatively high estimation errors observed with the KL-div loss, classification accuracy remains above 99%, highlighting the model's robustness to minor hyperparameter variations.

Finally, Fig. 6 shows the convergence of the different hyperparameters per modality. We notice that the hyperparameters associated with XRF modalities (Fe and Cu maps) converge faster and more precisely to their target values, due to their relative simplicity. In contrast, optical modalities (VIS, UVFC, IRFC) exhibit slightly larger estimation errors, although these remain small enough to have no significant impact on model performance.

5.2.2. Sensitivity analysis to noise and dataset size

While the theoretical convergence of the estimated hyperparameters to their target values is assessed using a noise-free synthetic dataset, we further evaluate classification performance under various noise conditions. Specifically, Gaussian noise is added to the synthetic features of each modality, with standard deviations expressed as percentages of each feature's dynamic range. The noise levels considered are 0%, 2%, 5%, and 10%. It is worth noting that the 10% noise level is already quite substantial, especially given that, in real-world experiments, the analysis is conducted at the superpixel level, where each feature vector aggregates information from multiple pixels. As a result, the pixel-level noise standard deviation is approximately $\sqrt{N_i}$ times larger than that at the superpixel level, where N_i denotes the number of pixels per superpixel (typically a few dozen). In addition, to assess the model's robustness more thoroughly, this evaluation is repeated for multiple dataset sizes, ranging from 10 K to 10M samples. The batch size was scaled with dataset size primarily to accelerate training: batch size $bs \in \{128, 512, 1024, 2048\}$ for datasets of size $ds \in \{10K, 100K, 1M, 10M\}$, respectively. Table 5 reports the mean and standard deviation of the classification accuracy, averaged across dataset sizes, for each noise level.

Table 6

Classification performance on the full mock-up dataset using the two supervised loss functions. Metrics include precision, recall, and F1-score for each class, as well as macro and weighted averages and overall accuracy.

Class	$\mathcal{L}_{\text{sup-1}}$			$\mathcal{L}_{\text{sup-2}}$		
	Prec.	Rec.	F1	Prec.	Rec.	F1
ω_{pb}	98	91	95	98	93	96
ω_{az}	99	85	91	99	92	95
ω_{ind}	78	78	78	81	80	81
ω_{\emptyset}	73	89	80	80	89	84
Macro Avg	87	86	86	89	89	89
Weighted Avg	89	87	87	91	90	90
Accuracy	87			90		

Table 7

Comparison of classification performance on the full mock-up dataset and the subset of mixtures containing at least 50% blue pigments. Metrics include precision, recall, and F1-score for each class, as well as macro and weighted averages and overall accuracy.

Class	Full Dataset			Pigment-Dominant Subset		
	Prec.	Rec.	F1	Prec.	Rec.	F1
ω_{pb}	98	93	96	99	99	99
ω_{az}	99	92	95	96	98	97
ω_{ind}	81	80	81	99	88	93
ω_{\emptyset}	80	89	84	89	95	92
Macro Avg	89	89	89	95	95	95
Weighted Avg	91	90	90	95	95	95
Accuracy	90			95		

The results show that classification performance remains highly stable with respect to dataset size, while accuracy gradually degrades as noise increases. Notably, the model maintains an accuracy above 87% up to a 5% noise level, which we consider satisfactory for this illustrative setting.

Furthermore, the low standard deviation of accuracy values across dataset sizes confirms that the training set size and the batch size have only a marginal influence on classification performance in this setup. This suggests that our approach is not particularly data-hungry and is capable of achieving competitive performance even with limited training samples.

5.3. Results on the mock-up dataset

This section presents the classification results obtained on the mock-up dataset. The comparative performance of $\mathcal{L}_{\text{sup-1}}$ and $\mathcal{L}_{\text{sup-2}}$ introduced in Section 4.4 is first discussed, followed by an evaluation and interpretation of classification performance. Next, we investigate the contribution of multimodality and optimization through an ablation study, and conclude with a comparative evaluation against SOTA evidential classifiers.

5.3.1. Ablation study: loss function comparison

Table 6 reports the classification metrics obtained using both loss functions on the full mock-up dataset. Overall, the two approaches produce comparable results, with $\mathcal{L}_{\text{sup-2}}$ yielding slightly better macro-average scores and overall accuracy. This suggests that the simplified loss, despite its compact formulation, remains effective at achieving strong classification performance. Based on these results, all subsequent analyses are conducted using $\mathcal{L}_{\text{sup-2}}$.

5.3.2. Analysis of classification behavior in full and pigment-dominant datasets

Detecting pigments in highly diluted mixtures—whether due to low pigment proportion or overall low concentration—can be particularly challenging for both optical and chemical modalities. To assess the impact of such detection difficulty on classification performance, we compare results obtained on the full mock-up dataset with those obtained on a *pigment-dominant* subset.

Table 7 presents the classification performance obtained on the full mock-up dataset as well as on a restricted version of it, which includes only mixtures containing at least 50% of blue pigments. Precision, recall, and F1-score are reported for each class, along with macro and weighted averages and overall accuracy. Class-level results are interpreted based on two parameters: the tinting

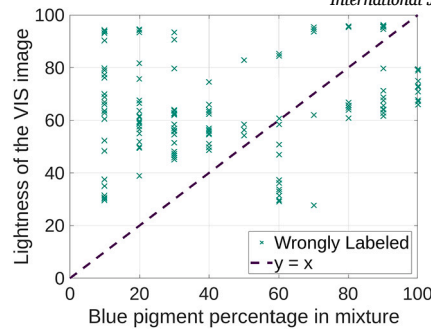


Fig. 7. Distribution of wrongly labeled samples in terms of lightness and blue pigment percentage in mixtures.

strength⁴ and the presence of a chemical indicator. The tinting strength of a pigment determines how easily it can be detected by optical modalities, as pigments with higher tinting strength visually dominate the mixtures. In contrast, the presence of a chemical indicator allows the XRF maps, which are generally more reliable than optical modalities, to contribute to the pigment identification.

Finally, classification challenges arise primarily for mixtures containing low proportions of blue pigments, particularly when the mixture is highly diluted, resulting in high lightness and reduced blue pigment concentration, complicating accurate identification, as illustrated in Fig. 7.

- Prussian Blue (ω_{pb}) shows balanced and robust performance across all metrics, with a 96% F1-score on the full dataset and a notable improvement to 99% on the pigment-dominant subset. This indicates that the model is able to consistently identify this pigment, regardless of its proportion in the mixtures.
- Azurite (ω_{az}) achieves the highest precision (99%) on the full dataset due to its high copper content that makes it clearly detectable using XRF. However, its recall is lower (92%) due to its weak tinting strength, which makes it difficult for optical imaging modalities to detect it in more diluted mixtures—leading to misclassifications as ω_{\emptyset} . In the pigment-dominant subset, recall rises to 98%, confirming that the model struggles to detect Azurite when its proportion in the mixture is low.
- Indigo (ω_{ind}) showed balanced precision (81%) and recall (80%), resulting in a F1-score of 81% on the full dataset. The slightly lower performance compared to other classes stems from the lack of a chemical indicator. Since Indigo is identified exclusively through optical modalities, its detection becomes more difficult in mixed or diluted samples. In the pigment-dominant subset, the F1-score improves to 93%, which is consistent with the fact that optical cues increase with pigment concentration.
- No Blue (ω_{\emptyset}) has the lowest precision (80%) on the full dataset, reflecting misclassifications of diluted samples of blue pigments. Nonetheless, its recall remains high (89%), showing the model's tendency to conservatively assign this class under uncertainty. On the pigment-dominant subset, precision increases to 89%.

These results demonstrate that classification errors in the full dataset are primarily due to challenging samples, particularly those with low pigment proportions or high lightness values. As shown in Fig. 7, misclassified samples are concentrated in regions characterized by low blue pigment content and high lightness, underscoring the role of mixture dilution in reducing detection reliability.

Finally, by examining the optimized discounting coefficients, we have observed that the estimated reliabilities for the Fe and Cu elements lie between 0.8 and 1. This is consistent with expectations, as these elements reliably indicate the presence of specific pigments (Prussian blue and azurite in our case). Conversely, for the optical data, the learned reliability coefficients exceed 0.7 only for indigo, which can be explained by the absence of a discriminative chemical signal for this pigment. For azurite in particular, the estimated reliability is notably lower, which is consistent with the qualitative observation that its characteristic color fades rapidly in optical modalities, as illustrated in Fig. 5 (second row).

5.3.3. Qualitative evaluation of the impact of multimodality

Fig. 8 illustrates a qualitative comparison highlighting the differences in pigment classification obtained when using optical modalities alone, XRF modalities alone, and multimodal fusion combining all modalities. Specifically, the figure depicts classification outcomes on mock-up samples representing each class. Due to Azurite's low tinting strength, the model relying solely on optical modalities struggles to accurately distinguish between ω_{az} and ω_{\emptyset} , leading to misclassification. For instance, Azurite is erroneously detected as present in samples where it should be labeled as ω_{\emptyset} . Conversely, the model based exclusively on XRF maps (Fe and Cu) mistakenly detects Indigo (ω_{ind}) in background areas. This misclassification occurs because Indigo lacks a chemical indicator, making it indistinguishable from the empty class using XRF modalities alone. Ultimately, the model that integrates both optical and XRF modalities addresses these individual shortcomings. By leveraging complementary information, this multimodal fusion accurately discriminates between the pigment classes and ω_{\emptyset} , significantly improving overall classification accuracy.

⁴ Tinting strength is the relative ability of a pigment to influence the color of a mixture.

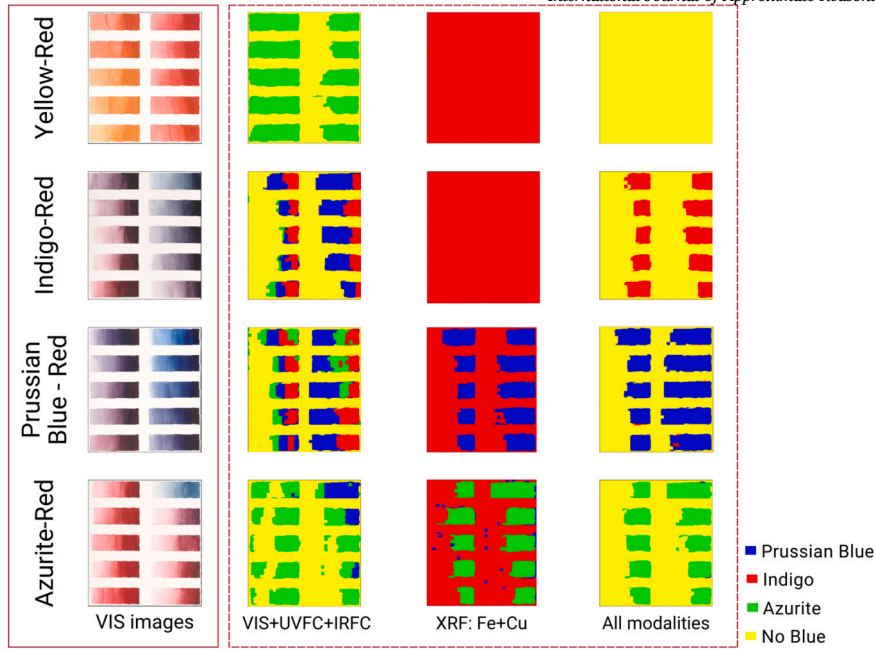


Fig. 8. Qualitative comparison of pigment classification results using optical modalities alone, XRF modalities alone, and multimodal fusion combining all modalities.

Table 8

Ablation study on the mock-up dataset (all values in percent, rounded to nearest integer). Best values in **bold**, second-best underlined.

Modality	Optimization	Accuracy	ω_{pb}	ω_{az}	ω_{ind}	ω_{\emptyset}
Precision / Recall						
XRF+Optical	No	49	62 / 42	50 / 11	43 / 83	47 / 77
Optical	Yes	60	91 / 71	60 / 18	<u>51</u> / 83	<u>49</u> / <u>78</u>
XRF	Yes	<u>63</u>	<u>97</u> / <u>89</u>	100 / <u>77</u>	30 / 46	0 / 0
XRF+Optical (ours)	Yes	90	98 / 93	<u>99</u> / 93	80 / <u>81</u>	81 / 89

5.3.4. Impact of multimodality and optimization

In this section, to better understand the individual contributions of the different components in our method, we conducted an ablation study focusing on the impact of multimodality and hyperparameter optimization.

We report in Table 8 comparative results that highlight the impact of multimodality and hyperparameter optimization in our approach. Specifically, we compare performance using a single type of modality (XRF or optical imaging) versus both, as well as the results before and after optimization.

According to these results, we clearly observe that optical modalities (i.e., VIS, UVFC, and IRFC) are critical for identifying classes such as ω_{ind} and ω_{\emptyset} , which cannot be reliably characterized by specific chemical elements. In contrast, elemental modalities (in this case Fe and Cu maps) are essential for detecting classes like ω_{pb} and ω_{az} . This confirms that only the full combination of all five modalities enables robust classification across all pigment classes.

Furthermore, while the overall accuracy of the multimodal setup without optimization reflects the benefit of qualitative constraints (as a random initialization without prior knowledge would yield approximately 25% accuracy in a four-class setting), the optimization step leads to an improvement in accuracy of up to 50%. This clearly underscores the added value of the neural-network-based optimization integrated into our fusion framework.

5.3.5. Comparison to SOTA evidential neural networks

In this section, we present a comparative analysis between our approach and two existing evidential classifiers. The first is the Evidential Neural Network (ENN) classifier [14]. The second is a model introduced in [28], inspired by [16], which interprets a Radial Basis Function (RBF) network as a combination of simple mass functions via Dempster's rule. We extended this model to support multiclass classification. For clarity, we refer to this method as the Evidential Radial Basis Function Network (ERBFN) classifier, and to our own model as NN-EMF (Table 9).

Our approach outperforms both evidential classifiers by at least 6% in overall accuracy, which we attribute to the integration of expert knowledge through qualitative constraints in the EMF representation. More importantly, the main advantage of our method lies in its ability to generate interpretable mass functions for each source, due to the structure imposed by domain-informed constraints.

Table 9

Comparison with SOTA evidential neural networks on the mock-up dataset (all values in percent, rounded to nearest integer). Best values in **bold**, second-best underlined.

Modality	Accuracy	ω_{ph}	ω_{az}	ω_{ind}	ω_{θ}
Precision / Recall					
ENN-classifier	82	93 / 97	82 / 74	90 / 85	67 / 72
ERBFN-classifier	<u>84</u>	<u>95</u> / <u>95</u>	<u>86</u> / <u>74</u>	92 / <u>85</u>	<u>68</u> / <u>80</u>
NN-EMF (ours)	90	98 / 93	99 / 93	80 / 81	81 / 89

The EMF hyperparameters correspond to easily interpretable quantities (e.g., source reliability, threshold values for pigment-related features), making the behavior of the fusion module transparent and interpretable for cultural heritage experts (in the context of our application).

6. Conclusion

This paper presents a general framework for assigning mass functions in multimodal settings, grounded in BFT. The proposed approach systematically integrates domain knowledge through parametric evidence modeling and refines mass assignments via a Constrained-NN optimized with task-specific loss functions. This enables the construction of reliable and adaptive belief representations, even under uncertain or variable data conditions.

From an application perspective, we presented one of the first multimodal pigment identification systems that effectively handles ambiguities across sources and demonstrates robustness to noise and variability. Experiments on synthetic and mock-up datasets confirmed the framework's ability to learn consistent evidence representations and achieve strong classification performance.

Open challenges remain. In particular, incorporating additional data modalities such as hyperspectral imaging raises the issue of managing high-dimensional feature spaces without compromising model interpretability. Furthermore, a deeper investigation into how BFT-based measures can be combined to design loss functions that reduce reliance on large labeled datasets would be interesting to improve generalization in limited-data scenarios.

CRediT authorship contribution statement

Sofiane Daimellah: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Sylvie Le Hégarat-Masle:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Clotilde Boust:** Writing – review & editing, Validation, Supervision, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. List of all abbreviations (Tables 10 and 11)

Table 10

List of abbreviations.

Abbreviation	Full Term
CH	Cultural Heritage
XRF	X-ray Fluorescence
VIS	Visible Imaging
UVFC	Ultraviolet False Color
IRFC	Infrared False Color
EMF	Evidence Mapping Function
BFT	Belief Function Theory
NN	Neural Network
MLP	Multi-Layer Perceptron
CNN	Convolutional Neural Network
FP	False Positive
FN	False Negative
CCE	Categorical Cross-Entropy

Table 10 (continued)

Abbreviation	Full Term
MAE	Mean Absolute Error
MSE	Mean Squared Error
KL-div	Kullback-Leibler Divergence
ENN	Evidential Neural Network
ERBFN	Evidential Radial Basis Function Network

Table 11

Key symbols and notations.

Notation	Description
$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \omega_\emptyset\}$	Frame of discernment containing the n pigments associated with a color c ; ω_\emptyset represents the absence of any pigment of color c .
$\Omega^* = \Omega \setminus \{\omega_\emptyset\}$	Set of all possible pigments (excluding the null hypothesis).
$\Theta_{mod} = \{\theta_{mod,1}, \dots, \theta_{mod,k}, \omega_\emptyset\}$	Modality-specific coarsening of Ω ; each $\theta_{mod,j}$ is a subset of pigments sharing similar characteristics in modality mod .
el	One XRF modality, referring to a specific chemical element (e.g., Fe, Pb).
opt	One optical modality (VIS, UVFC, or IRFC).
c_{el}	Pixel intensity in an XRF map representing the concentration of element el .
l_{opt}	Lightness component in optical modality opt .
h_{opt}	Hue angle in optical modality opt .
$\underline{\lambda}_{opt}, \bar{\lambda}_{opt}$	EMF hyperparameters defining lower and upper transition thresholds in modality opt .
α_{mod}	Discounting coefficient for the mass function of modality mod .
$\sigma_{opt,i}$	EMF hyperparameter defining the hue angle tolerance for pigments in $\theta_{opt,i}$.
$\gamma_{opt}, \bar{\gamma}_{opt}$	EMF hyperparameter controlling the uncertainty width zone in modality opt .
$\underline{\lambda}_{el}, \bar{\lambda}_{el}$	Transition thresholds in the l_{vis} space for element el .
$\underline{\rho}_{el}, \bar{\rho}_{el}$	Transition thresholds in the c_{el} space, obtained by projecting $\underline{\lambda}_{el}, \bar{\lambda}_{el}$ via r_{el}^{ph} .
r_{el}^{ph}	Correlation between c_{el} and l_{vis} for a given photograph ph .
m_{mod}	Mass function describing evidence from modality mod .

Appendix B. Case of Dubois-Prade rule associativity

In detection problems, the observations of a signal/response are usually associated with the presence of a detected/identified target. These observations are thus *Positives* measurements, either *true* when they actually correspond to an actual target, and *false* otherwise. However, the absence of signal/response, i.e. *Negatives* measurements, may also be exploited to support the (*true* or *false*) absence of a target.

In the following, the Positive and the Negative hypotheses in Ω discernment frame are denoted by H_i and H'_j , respectively. Then, we define:

$$\Omega = \{H_1, H_2, \dots, H_n, H'_{n+1}, \dots, H'_{n+n'}\}, \quad (\text{B.1})$$

$$\Omega^* = \{H_1, H_2, \dots, H_n\}, \quad (\text{B.2})$$

$$\Omega^\dagger = \{H'_{n+1}, \dots, H'_{n+n'}\} = \Omega \setminus \Omega^*. \quad (\text{B.3})$$

For example, in the considered pigment detection application, $n' = 1$, $\Omega^\dagger = \{\omega_\emptyset\}$ represents the absence of pigment, and, $\forall i \in \llbracket 1, n \rrbracket$, H_i represents either the pigments or distinct subsets of indistinguishable subsets of pigments, i.e. $\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \Rightarrow H_i \cap H_j = \emptyset$.

From Eq. (B.2), we also define: $\forall i \in \llbracket 1, n \rrbracket$,

$$A_i = \Omega^* \setminus \{H_i\}, \quad (\text{B.4})$$

which represents the alternative pigments to H_i . Notably, considering a hypothesis alongside its complementary (here in Ω^*) is a common strategy, as exemplified by the “one-versus-all” strategy [51] used for multi-class SVM. In the following, we refer to as “1-vs-all” focal sets those that have as possible focal elements the singleton H_i , sets of the form $A_i \cup B$ with $B \in 2^{\Omega \setminus A_i}$, or both:

$$\mathcal{F}_{m_i} \subseteq \mathcal{F}_i = \{H_i\} \cup \{A_i \cup B, B \in 2^{\Omega \setminus A_i}\}. \quad (\text{B.5})$$

Definition 1 (*Consistent focal sets*). A pair of “1-vs-all” focal sets \mathcal{F}_{m_i} and \mathcal{F}_{m_j} is said *consistent* with respect to discernment frame $\Omega = \Omega^* \cup \Omega^\dagger$ if there is a one-to-one relationship v_{ij} between \mathcal{F}_{m_i} and \mathcal{F}_{m_j} such that:

1. $v_{ij}(\emptyset) = \emptyset$,
2. $v_{ij}(H_i) = H_j$,
3. $v_{ij}(A_i) = A_j$,
4. $\forall (H, A, B) \in \{\emptyset, H_i\} \times \{\emptyset, A_i\} \times 2^{\Omega^\dagger}, v_{ij}(H \cup A \cup B) = v_{ij}(H) \cup v_{ij}(A) \cup B$.

This corresponds to replacing any occurrence of index i in a focal element involving H_i or A_i with an equivalent focal element using index j .

Definition 2 (Consistent set of focal sets). A set of “1-vs-all” focal sets $S = \{F_{m_i}, i \in \llbracket 1, n \rrbracket\}$ is said *consistent* with respect to discernment frame $\Omega = \Omega^* \cup \Omega^\dagger$ if, for any pair of focal sets $(F_{m_i}, F_{m_j}) \in S^2$, F_{m_i} and F_{m_j} are *consistent* according to Definition 1.

Proposition 1. Let a discernment frame $\Omega = \Omega^* \cup \Omega^\dagger$ (cf. Eq. (B.1)), for any consistent set of focal sets (cf. Definition 2), then Dubois-Prade’s combination rule (3) is associative:

$$\forall (i, j, k) \in \llbracket 1, n \rrbracket^3, (m_i \star m_j) \star m_k = m_i \star (m_j \star m_k).$$

Proof. Let us consider three BBAs, m_1 , m_2 , and m_3 , each having “1-vs-all” focal sets as defined in Eq. (B.5), denoted respectively by F_{m_1} , F_{m_2} , and F_{m_3} . For brevity, we denote the product $m_1(B_1)m_2(B_2)m_3(B_3)$ by $\Pi_{1,2,3}$, and express the Dubois-Prade combination rule as:

$$\begin{aligned} \forall A \in 2^\Omega \setminus \{\emptyset\}, (m_1 \star m_2) \star m_3(A) &= \sum_{B_3 \cap B = A} \left(\sum_{B_1 \cap B_2 = B} \Pi_{1,2,3} + \sum_{\substack{B_1 \cap B_2 = \emptyset \\ B_1 \cup B_2 = B}} \Pi_{1,2,3} \right) + \sum_{\substack{B_3 \cap B = \emptyset \\ B_3 \cup B = A}} \left(\sum_{B_1 \cap B_2 = B} \Pi_{1,2,3} + \sum_{\substack{B_1 \cap B_2 = \emptyset \\ B_1 \cup B_2 = B}} \Pi_{1,2,3} \right) \\ &= \sum_{B_1 \cap B_2 \cap B_3 = A} \Pi_{1,2,3} + \sum_{\substack{B_1 \cap B_2 = \emptyset \\ (B_1 \cup B_2) \cap B_3 = A}} \Pi_{1,2,3} + \sum_{\substack{B_1 \cap B_2 \neq \emptyset \\ (B_1 \cap B_2) \cap B_3 = \emptyset \\ (B_1 \cap B_2) \cup B_3 = A}} \Pi_{1,2,3} + \sum_{\substack{B_1 \cap B_2 = \emptyset \\ (B_1 \cup B_2) \cap B_3 = \emptyset \\ (B_1 \cup B_2) \cup B_3 = A}} \Pi_{1,2,3}. \end{aligned}$$

The first (pure conjunctive) term is known to be associative (conjunctive combination rule). Furthermore, the 1-vs-all structure ensures that the additional (disjunctive) terms collapse in a symmetric way. Thanks to the identity $\sum_{B \in 2^{\Omega \setminus A_i}} m_i(A_i \cup B) = 1 - m_i(\{H_i\})$, the possible focal elements in the combined mass associated with each one of the three remaining terms are presented as follows:

1. $\sum_{\substack{B_1 \cap B_2 = \emptyset \\ (B_1 \cup B_2) \cap B_3 = A}} \Pi_{1,2,3}$:
• $\{H_1, H_2\}$ with mass $m_1(\{H_1\})m_2(\{H_2\})(1 - m_3(\{H_3\}))$.
2. $\sum_{\substack{B_1 \cap B_2 \neq \emptyset \\ (B_1 \cap B_2) \cap B_3 = \emptyset \\ (B_1 \cap B_2) \cup B_3 = A}} \Pi_{1,2,3}$:
• $\{H_1, H_3\}$ with mass $m_1(\{H_1\})(1 - m_2(\{H_2\}))m_3(\{H_3\})$.
• $\{H_2, H_3\}$ with mass $(1 - m_1(\{H_1\}))m_2(\{H_2\})m_3(\{H_3\})$.
3. $\sum_{\substack{B_1 \cap B_2 = \emptyset \\ (B_1 \cup B_2) \cap B_3 = \emptyset \\ (B_1 \cup B_2) \cup B_3 = A}} \Pi_{1,2,3}$:
• $\{H_1, H_2, H_3\}$ with mass $m_1(\{H_1\})m_2(\{H_2\})m_3(\{H_3\})$.

After grouping these extra terms, one obtains an expression invariant under index permutation, thereby proving the associativity. \square

Data availability

Data will be made available on request.

References

- [1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
- [2] Viacheslav I. Adamchuk, R.A. Viscarra Rossel, Kenneth A. Sudduth, Peter Schulze Lammers, et al., Sensor Fusion for Precision Agriculture, Sensor Fusion-Foundation and Applications, vol. 13, 2011, pp. 27–40.
- [3] Mary B. Alatise, Gerhard P. Hancke, A review on challenges of autonomous mobile robot and sensor fusion methods, IEEE Access 8 (2020) 39830–39846.
- [4] Barbara H. Berrie, Rethinking the history of artists’ pigments through chemical analysis, Annu. Rev. Anal. Chem. 5 (1) (2012) 441–459.
- [5] A. Jizhen Cai, B. Clotilde Boust, C. Alamin Mansouri, An expert-inspired multimodal methodology for pigment identification in art paintings, in: 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2023, pp. 1–5.
- [6] G. Capobianco, Lucilla Pronti, E. Gorga, M. Romani, M. Cestelli-Guidi, Silvia Serranti, G. Bonifazi, Methodological approach for the automatic discrimination of pictorial materials using fused hyperspectral imaging data from the visible to mid-infrared range coupled with machine learning methods, Spectrochim. Acta, Part A, Mol. Biomol. Spectrosc. 304 (2024) 123412.
- [7] Chein-I. Chang, Spectral Information Divergence for Hyperspectral Image Analysis, IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS’99 (Cat. No. 99CH36293), vol. 1, IEEE, 1999, pp. 509–511.
- [8] Giacomo Chiari, David Scott, Pigment analysis: potentialities and problems, Period. Mineral. 73 (3) (2004) 227–237.
- [9] D.C. Creagh, D.A. Bradley, X-Ray Fluorescence Applications for the Study and Conservation of Cultural Heritage, Radiation in Art and Archeometry, 2000, p. 285.

- [10] Ana Luísa de Queiroz Baddini, Jorge Luiz Victor de Paula Santos, Raquel Reiner Tavares, Leticia Silva de Paula, Hiram da Costa, Araújo Filho, Renato P. Freitas, Pls-da and data fusion of visible reflectance, xrf and ftir spectroscopy in the classification of mixed historical pigments, *Spectrochim. Acta, Part A, Mol. Biomol. Spectrosc.* 265 (2022) 120384.
- [11] John K. Delaney, Jason G. Zeibel, Mathieu Thoury, R.O.Y. Littleton, Michael Palmer, Kathryn M. Morales, E. René de La Rie, A.N.N. Hoenigswald, Visible and infrared imaging spectroscopy of picasso's harlequin musician: mapping and identification of artist materials in situ, *Appl. Spectrosc.* 64 (6) (2010) 584–594.
- [12] Arthur P. Dempster, Upper and lower probability inferences based on a sample from a finite univariate population, *Biometrika* 54 (3–4) (1967) 515–528.
- [13] Xinyang Deng, Qi Liu, Yong Deng, Sankaran Mahadevan, An improved method to construct basic probability assignment based on the confusion matrix for classification problem, *Inf. Sci.* 340 (2016) 250–261.
- [14] Thierry Denoeux, A neural network classifier based on Dempster-Shafer theory, *IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum.* 30 (2) (2000) 131–150.
- [15] Thierry Denoeux, Decision-making with belief functions: a review, *Int. J. Approx. Reason.* 109 (2019) 87–110.
- [16] Thierry Denoeux, Logistic regression, neural networks and Dempster-Shafer theory: a new perspective, *Knowl.-Based Syst.* 176 (2019) 54–67.
- [17] Thierry Denoeux, Nn-evclus: neural network-based evidential clustering, *Inf. Sci.* 572 (2021) 297–330.
- [18] Thierry Denoeux, Orakanya Kanjanatarakul, Evidential clustering: a review, in: *International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*, Springer, 2016, pp. 24–35.
- [19] Thierry Denoeux, Orakanya Kanjanatarakul, Songsak Sriboonchitta, Ek-nnclus: a clustering procedure based on the evidential k-nearest neighbor rule, *Knowl.-Based Syst.* 88 (2015) 57–69.
- [20] Thierry Denoeux, M-H. Masson, Evclus: evidential clustering of proximity data, *IEEE Trans. Syst. Man Cybern., Part B, Cybern.* 34 (1) (2004) 95–109.
- [21] Sébastien Destercke, Thomas Burger, Toward an axiomatic definition of conflict between belief functions, *IEEE Trans. Cybern.* 43 (2) (2013) 585–596.
- [22] Didier Dubois, Henri Prade, A note on measures of specificity for fuzzy sets, *Int. J. Gen. Syst.* 10 (4) (1985) 279–283.
- [23] Didier Dubois, Henri Prade, Consonant approximations of belief functions, *Int. J. Approx. Reason.* 4 (5–6) (1990) 419–449.
- [24] Liguó Fei, Jun Xia, Yuqiang Feng, Luning Liu, A novel method to determine basic probability assignment in Dempster-Shafer theory and its application in multi-sensor information fusion, *Int. J. Distrib. Sens. Netw.* 15 (7) (2019) 1550147719865876.
- [25] Arthur Gestels, Francesca Gabrieli, Thomas De Kerf, Frederik Vanmeert, Hernan Fernández García, John Delaney, Koen Janssens, Gunther Steenackers, Steve Vanlanduit, High-resolution compound-specific mapping in works of art via data fusion of ma-xrpd with hyperspectral data (part 1: method evaluation), *Talanta* 280 (2024) 126731.
- [26] Bartosz Grabowski, Wojciech Masarczyk, Przemysław Głomb, Agata Mendys, Automatic pigment identification from hyperspectral data, *J. Cult. Heritage* 31 (2018) 1–12.
- [27] Ulrich Höhle, Entropy with respect to plausibility measures, in: *Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic*, Paris, 1982, 1982.
- [28] Ling Huang, Su Ruan, Pierre Decazes, Thierry Denoeux, Lymphoma segmentation from 3d pet-ct images using a deep evidential network, *Int. J. Approx. Reason.* 149 (2022) 39–60.
- [29] Ariane Isler-de Jongh, The origins of colour photography: scientific, technical and artistic interactions, *Hist. Photogr.* 18 (2) (1994) 111–119.
- [30] Wen Jiang, Boya Wei, Xiyun Qin, Jun Zhan, Yongchuan Tang, Sensor data fusion based on a new conflict measure, *Math. Probl. Eng.* 2016 (1) (2016) 5769061.
- [31] Wen Jiang, Jun Zhan, Deyun Zhou, Xin Li, A method to determine generalized basic probability assignment in the open world, *Math. Probl. Eng.* 2016 (1) (2016) 3878634.
- [32] Wen Jiang, Miaoyan Zhuang, Chunhe Xie, A reliability-based method to sensor data fusion, *Sensors* 17 (7) (2017) 1575.
- [33] Cerys Jones, Nathan S. Daly, Catherine Higgitt, Miguel R.D. Rodrigues, Neural network-based classification of x-ray fluorescence spectra of artists' pigments: an approach leveraging a synthetic dataset created using the fundamental parameters method, *Herit. Sci.* 10 (1) (2022) 1–14.
- [34] Anne-Laure Jousset, Dominic Grenier, Éloi Bossé, A new distance between two bodies of evidence, *Inf. Fusion* 2 (2) (2001) 91–101.
- [35] Moshe Kam, Xiaoxun Zhu, Paul Kalata, Sensor fusion for mobile robot navigation, *Proc. IEEE* 85 (1) (1997) 108–119.
- [36] Tania Kleyhans, Catherine M. Schmidt Patterson, Kathryn A. Dooley, David W. Messinger, John K. Delaney, An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence, *Herit. Sci.* 8 (2020) 1–16.
- [37] Maria T. Lamata, Serafin Moral, Measures of entropy in the theory of evidence, *Int. J. Gen. Syst.* 14 (4) (1988) 297–305.
- [38] Eberhard H. Lehmann, Peter Vontobel, Eckhard Deschler-Erb, Marie Soares, Non-invasive studies of objects from cultural heritage, *Nucl. Instrum. Methods Phys. Res., Sect. A, Accel. Spectrom. Detect. Assoc. Equip.* 542 (1–3) (2005) 68–75.
- [39] Junwei Li, Baolin Xie, Yong Jin, Lin Zhou, Generating method and application of basic probability assignment based on interval number distance and model reliability, *Soft Comput.* 28 (3) (2024) 2353–2365.
- [40] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, Jun Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE Trans. Neural Netw. Learn. Syst.* 33 (12) (2021) 6999–7019.
- [41] Lingxi Liu, Tsveta Miteva, Giovanni Delnevo, Silvia Mirri, Philippe Walter, Laurence de Viguier, Emeline Pouyet, Neural networks for hyperspectral imaging of historical paintings: a practical review, *Sensors* 23 (5) (2023) 2419.
- [42] Yu Liu, Shuping Liu, Zengfu Wang, A general framework for image fusion based on multi-scale transform and sparse representation, *Inf. Fusion* 24 (2015) 147–164.
- [43] Tianshuo Ma, Fuyuan Xiao, An improved method to transform triangular fuzzy number into basic belief assignment in evidence theory, *IEEE Access* 7 (2019) 25308–25322.
- [44] Marie-Hélène Masson, Thierry Denoeux, Ecm: an evidential version of the fuzzy c-means algorithm, *Pattern Recognit.* 41 (4) (2008) 1384–1397.
- [45] Hung T. Nguyen, On entropy of random sets and possibility distributions, *Anal. Fuzzy Inf.* 1 (1987) 145–156.
- [46] Using Fuzzy Numbers, Determination of basic belief assignment using fuzzy numbers, *Adv. Appl. DSmT Inf. Fus.* 623 (2023) 1–6.
- [47] Antonio Plaza, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, et al., Recent advances in techniques for hyperspectral image processing, *Remote Sens. Environ.* 113 (2009) S110–S122.
- [48] Marius-Constantin Popescu, Valentina E. Balas, Liliana Perescu-Popescu, Nikos Mastorakis, Multilayer perceptron and neural networks, *WSEAS Trans. Circuits Syst.* 8 (7) (2009) 579–588.
- [49] Emeline Pouyet, Tsveta Miteva, Neda Rohani, Laurence de Viguier, Artificial intelligence for pigment classification task in the short-wave infrared range, *Sensors* 21 (18) (2021) 6150.
- [50] Bowen Qin, Fuyuan Xiao, An improved method to determine basic probability assignment with interval number and its application in classification, *Int. J. Distrib. Sens. Netw.* 15 (1) (2019) 1550147718820524.
- [51] Ryan Rifkin, Aldebaro Klautau, In defense of one-vs-all classification, *J. Mach. Learn. Res.* 5 (Jan 2004) 101–141.
- [52] Neda Rohani, Emeline Pouyet, Marc Walton, Oliver Cossairt, Aggelos K. Katsaggelos, Nonlinear unmixing of hyperspectral datasets for the study of painted works of art, *Angew. Chem.* 130 (34) (2018) 11076–11080.
- [53] Neda Rohani, Emeline Pouyet, Marc Walton, Oliver Cossairt, Aggelos K. Katsaggelos, Pigment unmixing of hyperspectral images of paintings using deep neural networks, in: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 3217–3221.
- [54] Neda Rohani, Johanna Salvant, Sara Bahaadini, Oliver Cossairt, Marc Walton, Aggelos Katsaggelos, Automatic pigment identification on roman Egyptian paintings by using sparse modeling of hyperspectral images, in: *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, pp. 2111–2115.
- [55] Mojtaba Safari, Ali Fatemi, Louis Archambault, Medfusiongan: multimodal medical image fusion using an unsupervised deep generative adversarial network, *BMC Med. Imaging* 23 (1) (2023) 203.

- [56] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
- [57] Nirmesh Sharma, Challenges faced by the bhotias for their livelihood and preservation of culture, *Int. J. Comp. Sociol. Anthropol.* 12 (2) (2020) 51–58.
- [58] Jacob Sherman, The theoretical derivation of fluorescent x-ray intensities from mixtures, *Spectrochim. Acta* 7 (1955) 283–306.
- [59] Ewen Smith, Geoffrey Dent, *Modern Raman Spectroscopy: a Practical Approach*, John Wiley & Sons, 2019.
- [60] V.A. Solé, E. Papillon, M. Cotte, Ph. Walter, J. Susini, A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra, *Spectrochim. Acta, Part B, At. Spectrosc.* 62 (1) (2007) 63–68.
- [61] Jana Striova, Alice Dal Fovo, Raffaella Fontana, Reflectance imaging spectroscopy in heritage science, *Riv. Nuovo Cimento* 43 (2020) 515–566.
- [62] Yongchuan Tang, Yonghao Zhou, Xiangxuan Ren, Yufei Sun, Yubo Huang, Deyun Zhou, A new basic probability assignment generation and combination method for conflict data fusion in the evidence theory, *Sci. Rep.* 13 (1) (2023) 8443.
- [63] D. Thickett, C.S. Cheung, H. Liang, J. Twyde, R. Gr Maev, D. Gavrilo, Using non-invasive non-destructive techniques to monitor cultural heritage objects, *Insight* 59 (5) (2017) 230–234.
- [64] Sebastian Thrun, Probabilistic robotics, *Commun. ACM* 45 (3) (2002) 52–57.
- [65] Bingjie Jenny Xu, Yunan Wu, Pengxiao Hao, Marc Vermeulen, Alicia McGeachy, Kate Smith, Katherine Eremin, Georgina Rayner, Giovanni Verri, Florian Willomitzer, et al., Can deep learning assist automatic identification of layered pigments from xrf data?, *J. Anal. At. Spectrom.* 37 (12) (2022) 2672–2682.
- [66] Ronald R. Yager, Entropy and specificity in a mathematical theory of evidence, in: *Classic Works of the Dempster-Shafer Theory of Belief Functions*, 2008, pp. 291–310.
- [67] Li Yang, Björn Kruse, Revised Kubelka–Munk theory. I. Theory and application, *JOSA A* 21 (10) (2004) 1933–1941.
- [68] Chao Zhang, Zichao Yang, Xiaodong He, Li Deng, Multimodal intelligence: representation learning, information fusion, and applications, *IEEE J. Sel. Top. Signal Process.* 14 (3) (2020) 478–493.
- [69] Chenwei Zhang, Yong Hu, Felix T.S. Chan, Rehan Sadiq, Yong Deng, A new method to determine basic probability assignment using core samples, *Knowl.-Based Syst.* 69 (2014) 140–149.
- [70] Jingfei Zhang, Yong Deng, A method to determine basic probability assignment in the open world and its application in data fusion and classification, *Appl. Intell.* 46 (4) (2017) 934–951.
- [71] Siqiang Zhong, Xingcheng Liu, A new method to determine basic probability assignment based on interval number, in: *2019 Computing, Communications and IoT Applications (ComComAp)*, IEEE, 2019, pp. 316–320.