Journal Pre-proofs

Review Article

Multi-source information fusion: Progress and future

Xinde LI, Fir DUNKIN, Jean DEZERT

PII: S1000-9361(23)00424-7
DOI: https://doi.org/10.1016/j.cja.2023.12.009
Reference: CJA 2874

To appear in: Chinese Journal of Aeronautics

Received Date: 14 July 2023
Accepted Date: 5 December 2023

Please cite this article as: X. LI, F. DUNKIN, J. DEZERT, Multi-source information fusion: Progress and future, Chinese Journal of Aeronautics (2023), doi: https://doi.org/10.1016/j.cja.2023.12.009

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Production and hosting by Elsevier Ltd. on behalf of Chinese Society of Aeronautics and Astronautics.
Multi-source information fusion: Progress and future

Xinde LI\textsuperscript{a,b,c,d,}\textsuperscript{*}, Fir DUNKIN\textsuperscript{a,b,c}, Jean DEZERT\textsuperscript{e}

\textsuperscript{a} School of Automation, Southeast University, Nanjing, 210008, China
\textsuperscript{b} Nanjing Center for Applied Mathematics, Nanjing, 211135, China
\textsuperscript{c} Key Laboratory of Measurement and Control of CSE, Ministry of Education, Nanjing, 210096, China
\textsuperscript{d} Southeast University Shenzhen Research Institute, Shenzhen, 518063, China
\textsuperscript{e} Information Modeling and Processing Department, ONERA, Palaiseau, 92320, France

KEYWORDS
Multi-sensor system; Information fusion; Artificial intelligence; Pattern recognition; Human-machine integration

Abstract Multi-Source Information Fusion (MSIF), as a comprehensive interdisciplinary field based on modern information technology, has gained significant research value and extensive application prospects in various domains, attracting high attention and interest from scholars, engineering experts, and practitioners worldwide. Despite achieving fruitful results in both theoretical and applied aspects over the past five decades, there remains a lack of comprehensive and systematic review articles that provide an overview of recent development in MSIF. In light of this, this paper aims to assist researchers and individuals interested in gaining a quick understanding of the relevant theoretical techniques and development trends in MSIF, which conducts a statistical analysis of academic reports and related application achievements in the field of MSIF over the past two decades, and provides a brief overview of the relevant theories, methodologies, and application domains, as well as key issues and challenges currently faced. Finally, an analysis and outlook on the future development directions of MSIF are presented.

1. Introduction

Humans utilize a “multi-sensor system” composed of eyes, ears, nose, tongue, and body to gather information from multiple sources when describing things (see Fig. 1). This information is then integrated in the brain through Multi-Source Information Fusion (MSIF) to arrive at a final decision. The reason for choosing this paradigm is that compared to single sensor measurement, MSIF has the potential to obtain higher precision information and stronger system survival ability, as well as the advantages of wide detection spatiotemporal range, low cost, light weight, and less duty cycle.\textsuperscript{1}

\textsuperscript{*}Corresponding author
Email address: xindeli@seu.edu.cn (Xinde LI)
Peer review under responsibility of Editorial Committee of CJA.

Meanwhile, information fusion plays a crucial role in the aviation field, which is reflected in improving flight safety, navigation
accuracy, onboard system performance, maintenance efficiency, and aviation management in all aspects, which helps to improve the efficiency, reliability, and service level of the aviation industry, and is one of the indispensable core technologies in the modern aviation field.

Specifically, information fusion can integrate multiple sources of data (e.g., radar, satellites, meteorology, communication, etc.) together, providing a more comprehensive understanding of the flight environment, enabling pilots to better understand the flight environment and respond to unexpected situations, thereby reducing accident risks; information fusion can integrate data from satellite navigation systems with other navigation information, improve the accuracy and controllability of flight paths, and reduce navigation errors and flight delays; information fusion integrates data from various onboard systems in the aircraft, providing better operation and monitoring functions; fault diagnosis based on information fusion can be used to integrate data from aircraft sensors and monitoring systems, achieve effective maintenance and health monitoring, reduce maintenance costs, reduce aircraft downtime, and improve flight reliability; information fusion is also crucial for aviation traffic management, which can help airlines and air traffic control departments better coordinate and manage flight plans, improve air traffic smoothness, and reduce congestion and delays.

1.1. Definition

In the late 1970s, terms such as “data fusion” or “sensor fusion” appeared in various publications and academic reports. However, there was no unified concept of information fusion and related technologies at that time, until the widespread use of “information fusion” in the 1990s. The emerging interdisciplinary field of MSIF has only gained a relatively unified academic term. Afterwards, the relevant theoretical technologies of MSIF have been rapidly developed (Fig. 23-14), and fruitful results have been achieved in military (e.g., military automation systems, strategic early warning and defense systems, multi target tracking and recognition, and damage effect assessment) and civilian (including remote sensing, smart medical diagnosis, e-commerce, wireless communication, navigation positioning, fault diagnosis, etc.) fields.

Despite nearly fifty years of development, the academic community and industry have yet to reach a consensus on the definition of MSIF. The Joint Directors of Laboratories (JDL) of the United States Military Organization has provided Definition 1 from a military application perspective. Based on this, Klein expands information fusion technology from single sources by defining and explaining data sources. Subsequently, Han et al. and Pan et al. summarized their research on JDL and provided Definition 2. Then, Khaleghi et al. proposed Definition 3 based on the viewpoint of Boström et al. Finally, based on the latest academic reports and research results, we describe MSIF as Definition 4.

Definition 1. **MSIF is a multi-level and multifaceted processing process that includes detecting, associating, combining, and estimating multi-source data to improve the accuracy of state and identity estimation, as well as timely and complete assessment of the importance of battlefield situation and threats.**

Definition 2. **MSIF, also known as multi-sensor fusion, is a multi-level and multifaceted process that includes the detection, association, combination, and estimation of multi-source data to improve the accuracy of information, state, and identity estimation, as well as a timely and complete assessment of the ultimate degree of target situation and threat.**

Definition 3. **MSIF is an effective method for automatically or semi-automatically transforming information from different sources and time points into representations that provide effective support for human or automated decision-making.**

Definition 4. **MSIF is a process of detecting, representing, mining, associating, synthesizing, reasoning, and predicting multi-source, multi-level, multi-dimensional, and granular data, information, and knowledge, which integrates modern information technologies, including Artificial Intelligence (AI), knowledge reasoning, data analysis, processing, etc., to obtain richer and more accurate consistent descriptions or decision results of the observed object.**

1.2. Background

1.2.1. Times

With the rapid development of technologies such as AI and the Internet of Things, an increasing number of smart sensors are changing human lifestyles and even social structures. Numerous multi-sensor systems targeting complex environments have significantly improved the efficiency of information acquisition, greatly facilitating the development of automated systems based on information interaction and processing. Global data is growing at an annual rate of 30%, and massive amounts of data are being created, collected, and processed. Therefore, how to extract reliable, accurate, and valuable information from big data has become one of the most important research topics today.

Currently, data appears in the form of multiple sources, not limited to a single source, in terms of storage, acquisition, and description. Different data sources often contain rich underlying knowledge. MSIF combines multiple information sources with spatial redundancy, temporal redundancy, or complementary information constraints, thus possessing the potential to fully utilize various information sources. Compared to raw sensor data, fused decisions (descriptions) have more consistent, abundant, and accurate information.

In this era, research on MSIF has significant temporal value and practical significance.

1.2.2. Strategic

On October 18, 2021, General Secretary Xi Jinping mentioned “promoting the deep integration of digital technology and the real economy” and “giving birth to new industries, new forms, and new models” during the 34th collective study of the Central Political Bureau of the Communist Party of China. Subsequently, the Ministry of Industry and Information Technology officially issued the “14th Five-Year Plan for the Deep Integration of Informatization and Industrialization” which comprehensively deployed the work focus of “integrating informatization and industrialization during the 14th Five-Year Plan period”, which marks the digitalization, digitization, and intelligentization as the main trends for the future development of China’s industry and signifies that the digital economy will usher in new development opportunities.

Therefore, the development of MSIF technology, with the main goal of solving multi-data synthesis decision-making, becomes essential in implementing the deep integration development of digital technology and sectors related to people’s livelihoods and intelligent applications. In addition, China’s 14th Five-Year Plan also puts forward strategic goals such as “accelerating national defense and military modernization” and “improving the quality and efficiency of national defense and military modernization.” MSIF technology stems from the digitalization, intelligentization, and modernization needs of command systems and military equipment systems in the
late 20th century of the United States, making the development of MSIF technology an essential path to enhance the quality of China’s national defense and military modernization.

In this strategic context, conducting research and development on the theoretical and technical aspects of MSIF technology, which was born to meet military needs and has been widely applied in civilian fields, holds significant practical value and strategic significance.

1.2.3. Academic

To meet the needs of processing multiple independent sonar signals for information fusion, the US military proposed a method for enemy submarine detection based on MSIF in the 1970s. Since then, MSIF has been widely applied in military systems such as C³I, Information Warfare (IW), and C⁴I, gradually becoming a comprehensive interdisciplinary field with the ability to break down disciplinary barriers and provide opportunities for collaboration, reference, and exchange among scholars in different research fields.

With the continuous convening of conferences such as the International Conference on Information Fusion (ICIF), the International Conference on Information and Data Fusion (ICIDF), the IEEE International Symposium on Information Theory (IEEE ISIT), and the International Society of Information Fusion (ISIF), researchers in relevant fields at home and abroad have been facilitated to promptly understand and grasp the latest technologies, theoretical developments, and achievements related to MSIF, which has also led to an increasing number of scholars paying attention to research in this direction, resulting in a vast number of achievements in the theory and applications of MSIF in recent years.

Taking the representative journal Information Fusion in the field of MSIF as an example, the number of publications has gradually increased from around 80 papers in 2018 to 239 papers in 2022. The impact factor has also steadily grown, from 10.716 to 18.511, which indicates that more and more scholars and practitioners have a strong interest in the theory and applications of MSIF.

In this academic background, it is of urgent practical significance and research value to sort out and summarize the technical theories and academic achievements of MSIF.

1.3. Motivation

We have compiled articles containing the keyword “information fusion” from the publications included in Web of Science from 2000 to 2022, as shown in Fig. 3. Information fusion is becoming a hot research direction from an interdisciplinary field.

In the past two decades, there has been a continuous increase in the number of reports on “information fusion” and with the advent of the big data era, there has been a rapid growth trend since 2017 (see Fig. 3). Clearly, the explosive growth of achievements in the fields of computer science and engineering has been the main driving force behind the rapid development of information fusion, because the rapid development of AI and pattern recognition technologies, represented by machine learning and deep learning, has significantly enriched the strategies and approaches of information fusion, enabling information fusion techniques and theoretical methods to have the ability to solve a wide range of engineering problems.

Benefiting from the assistance of computer technology in the
development of information fusion theories and techniques, information fusion technology has achieved fruitful results not only in the field of engineering but also in domains, e.g. communications, instrumentation, and automation, which has consistently achieved publication volumes exceeding one thousand papers in multiple research fields, which indicates that with the advent of the digital age, the demand for fusion processing of different types of data is increasing, and the demand for the development of information fusion technology is becoming increasingly strong.

Despite the fruitful research achievements and considerable attention from scholars in various research fields and engineering applications, there are still some issues for the current review articles on MSIF, early reporting, incomplete review, insufficient depth, etc. (Fig. 4)

![Fig. 4 Comparison with other reviews. The radius of each circle is positively correlated with the comprehensiveness of the corresponding review, with red indicating the work where MSIF has updated definitions, and green expressing that the theme has not provided new definitions. Simultaneously, the hollow circle represents that there is no reported content related to deep learning. More detailed information can be found in Table 1.](image)

Table 1 summarizes the statistical information of relevant reviews. 26–29,37–43 As shown in Table 1 and Fig. 4, the currently published reviews on MSIF can be roughly divided into two categories: comprehensive reviews that were published earlier (such as Ref. 26 in Jul. 2002, Ref. 27 in Jul. 2003, and Ref. 28 in Jan. 2013), and recently published work on a single subtopic (e.g., Ref. 39 focuses on MSIF in wearable sensor networks, Ref. 40 only reviews the application of rough set theory in MSIF, Ref. 29 focuses on MSIF in human behavior recognition, and Ref. 42 summarizes the application of MSIF in the field of information enhancement.).

In summary, considering the background of the big data era and the current academic status quo of the lack of novel and comprehensive review articles on MSIF, and in response to the call of the national “14th Five-Year Plan” for the development of information theory, we argue that it is necessary to comprehensively review the development status, application domains, key issues, and future directions of MSIF as a discipline over the past two decades. This work, therefore, conduct a comprehensive statistical analysis of the current theoretical and application achievements of MSIF, with the aim of helping relevant scholars and researchers quickly understand the current situation and development trends in this field, and facilitating researchers and practitioners in other fields to establish connections between their work and MSIF, thereby contributing to the promotion of the development of relevant theories, technologies, and methodologies to the best of our abilities.

**Note:** We mainly focus on the recent achievements and future development trends of MSIF, and Refs. 26–28 provide a brief overview of the origin and development of MSIF; thereby this article does not elaborate on the origin of MSIF.

### 1.4. Main content of this article

As shown in Fig. 5, the structure of this paper is given as follows: First, Section 1 provides a description of the definition and research background of MSIF, along with an introduction to the motivation behind this paper. Next, Section 2 elaborates on the architectures, models, theories, and implementation approaches of MSIF. Following that, Section 3 introduces and analyzes the main application areas of MSIF from both military and non-military perspectives. Then, in Section 4, key issues and challenges faced by MSIF are discussed from the aspects of data and applications. Future development directions are also anticipated based on the current development status. Finally, Section 5 provides a brief summary.

### 2. Architectures, models, and methodologies

On the technical level, MSIF draws inspiration from various fields, such as signal processing, information theory, statistical estimation and inference, and AI. Traditional fusion methods are primarily based on statistical inference and estimation theory, which have laid a solid theoretical foundation for MSIF. In recent years, the emergence of new methods based on technologies like AI and information theory has led to significant progress and development in the relevant techniques of MSIF. In this section, a concise description is provided regarding the architectures for MSIF, models of fusion systems, and the theories and methods of information fusion.

---

**Table 1 Related reviews.**

<table>
<thead>
<tr>
<th>Theme</th>
<th>Time</th>
<th>Ref.</th>
<th>Comprehensiveness</th>
<th>IDL</th>
<th>Def</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mod the met app</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref. 26</td>
<td>2002(Jul.)</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Ref. 27</td>
<td>2003(Jul.)</td>
<td>139</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Ref. 37</td>
<td>2009(Sep.)</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ref. 28</td>
<td>2013(Jan.)</td>
<td>197</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Ref. 38</td>
<td>2013(Aug.)</td>
<td>40</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ref. 39</td>
<td>2017(May.)</td>
<td>179</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ref. 40</td>
<td>2019(Aug.)</td>
<td>162</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Ref. 41</td>
<td>2021(Dec.)</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Ref. 42</td>
<td>2022(Feb.)</td>
<td>172</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ref. 43</td>
<td>2022(Feb.)</td>
<td>27</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Ref. 29</td>
<td>2022(Apr.)</td>
<td>310</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Proposed</td>
<td>2023</td>
<td>271</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

**Note:** In the table, “Time” is the time of publication, “Ref.” refers to the total number of references in the theme, “IDL” indicates whether deep learning is involved, “Def” refers to whether the author has summarized or updated the specific definition of MSIF based on the literature at that time, and the statistics in “Comprehensiveness” are the number of fusion models (“mod”), theories (“the”), methodologies (“met”), and application fields (“app”) detailed in the review.

---

![Fig. 5 Organizational structure and main content of this article.](image)
2.1. Architectures for MSIF

As a comprehensive discipline, MSIF encompasses not only the strategies for acquiring integrated decision-making from multiple data sources but also the theoretical methods for processing the fused data. Based on the level at which information sources are processed, MSIF methods can be categorized into three types: data-level fusion, feature-level fusion, and decision-level fusion (Fig. 6).

1. Data-level fusion. The lowest level of information fusion is data-level fusion (Fig. 6(a)). By inputting raw data captured from sensors, it is possible to obtain data output with high accuracy, reliability, and low noise. Data-level fusion is often employed in information fusion scenarios where the preservation of more original information is desired. Therefore, data-level fusion is commonly used in signal fusion, image fusion (also known as pixel fusion), and similar applications.

2. Feature-level fusion. In feature-level fusion (Fig. 6(b)), the input to the fusion model consists of data or features that have already been extracted. The output of the fusion algorithm is either refined features or higher-level data (i.e., decisions) in different forms. Compared to data-level fusion, feature-level fusion methods provide more detailed and comprehensive information, showcasing various characteristics of the data.

3. Decision-level fusion. To further integrate the generated information to reveal more comprehensive decision-making in tasks, the highest level of fusion known as decision-level fusion, is commonly employed. This type of fusion strategy not only requires local decisions derived from individual perspectives but also provides integrated decisions with global information. Therefore, decision-level fusion typically occurs before making final decisions. Compared to data-level fusion and feature-level fusion, decision-level fusion methods generate preliminary discriminative results and then fuse decisions from different types of data to obtain more comprehensive and accurate outcomes.

As shown in Fig. 6(d), data-level fusion can effectively preserve the original measurement information of the tested target, but due to the large amount of data, the real-time performance of fusion processing is poor, and can only process fused isomorphic data. Feature-level fusion can achieve information fusion of similar or heterogeneous sensors, with a small amount of data and high real-time algorithm, which can also process and fuse heterogeneous data through multi-modal data encoding, but information loss is caused to a certain extent, due to the compression of the original measurement information. Decision-level fusion can complete the information fusion task of similar or heterogeneous sensors, with small communication volume and strong anti-interference ability, which can easily fuse and analyze heterogeneous data, but the information loss is relatively serious.

2.2. Models of information fusion system

With the development of hardware devices and relevant instruments in various fields of computer science and engineering, the diversity of data information used to characterize the operational state of targets has been greatly enriched. MSIF, through the combined analysis of data from multiple homogeneous or heterogeneous sensors, has achieved fruitful results in numerous military and civilian domains, providing more accurate situational awareness descriptions. Due to the wide-ranging applications of MSIF, there is currently no universal data fusion implementation framework, and the industry and academia generally do not recommend using a specific or fixed data fusion architecture for various different applications.

The first fusion model was proposed by Bowman in 1980, known as the Bowman Df&Rm model. Later, in 1988, Luo et al. introduced an information fusion process based on multi-sensor data fusion, i.e., the Luo-Kay model. In the same year, Pau proposed a hierarchical architecture and model for knowledge-based data fusion. The most widely used information fusion framework today is the JDL model, which originated from the military domain in the early 1990s. Over the decades of development, both academia and industry have put forth numerous model frameworks for information fusion systems. To avoid redundancy and excessive length, this section focuses on introducing the four most representative models, namely the JDL model, Bowman Df&Rm model, Luo-Kay model, and Pau model (Fig. 7).

2.2.1. JDL model

The JDL model focuses on data fusion (correlation, filtering, and association), and the specific structure is depicted in Fig. 7(a), where data flows from data sources to the human-computer interface in a sequential manner through five stages: sources preprocessing (level-0), object refinement (level-1), situation refinement (level-2), threat refinement (level-3), and process refinement (level-4). The ability to utilize contextual information is crucial for transitioning from low-level to high-level information fusion. To enable the JDL model to incorporate contextual information, the Data Fusion Information Group (DFIG) model was proposed in 1998 as an important extension. The DFIG model introduces data fusion and resource management capabilities into the system, emphasizing the participation of users as observers, which helps to infer contextual information during the data fusion process. Subsequently, in 2004, researchers and practitioners re-evaluated the JDL model, particularly the DFIG-based structure. They identified shortcomings for example “the need for users to apply various skills (perception, rules (tasks), and cognition (knowledge) to accomplish different tasks, and many JDL model systems only meet specific task requirements”. Since the revision in 2004, researchers have...
focused on information management, advanced visualization, and data mining, as well as team, priority, and coordination tasks when designing fusion systems based on the DFIG model. These functionalities contribute to addressing the challenges of fusion and resource management within the DFIG model.

Two expansions in 1998 and 2004 have somewhat alleviated the issues of the JDL model, which put the emphasis on data (input/output) rather than fusion processing and the excessive constraints. However, the JDL model still faces limitations when applied to practical tasks. To address limitations, the Dasarathy model considers fusion systems from a software engineering perspective, treating them as data flows characterized by input/output and functions (processes). Dasarathy fusion paradigm is primarily used to make optimal decisions in multi-sensor target recognition and tracking environments, and aims to fuse data from a set of sensors in parallel and embeds it within a recursive system structure to enhance the reliability of fusion decisions. Similarly, addressing the issue of representing decision reliability, Goodman et al. proposed another fusion model based on the concept of random sets, which combines decision uncertainty with the decision itself and provides a fully general scheme for representing uncertainty.

2.2.2. Bowman Df&Rm model

The Bowman Df&Rm (Data Fusion & Resource Management) model, proposed by Bowman in 1980, is a generic data fusion architecture for addressing multi-sensor, multi-target identification, and tracking problems. While the JDL model has demonstrated effectiveness in many data fusion applications, Bowman assumed that JDL model had limited impact on the architectural development of practical systems. Consequently, Bowman introduced the concept of a data fusion hierarchy tree to partition fusion problems into nodes, where each node conceptually involves functions such as data linking, estimation, and association.

In the Bowman Df&Rm model, a data processing approach based on correlation assumptions and prior information is employed to assess the relative uncertainty impact of uncertain, unknown targets on the data and decision outcomes. A structure comprising hypothesis generation and evaluation feedback (hypothesis validation) is proposed, as shown in Fig. 7(b). Furthermore, the Bowman Df&Rm model exhibits duality between estimation and control, wherein the data fusion and resource management systems can be realized through the combination and management of interactions between network nodes, which makes the interpretation of different levels of data types, sources, models, and conclusions distinct within the Bowman Df&Rm architecture. In the context of the Bowman Df&Rm model, the process refinement functionality, which is considered one of the component resource management functions, corresponds to the fourth level in the JDL model revised in 2004.

2.2.3. Luo-Kay model

The Luo-Kay model, proposed by Luo and Kay in 1988, presents a generic data fusion structure based on the integration of multiple sensors. In the Luo-Kay model, data from multiple sources is combined in a hierarchical manner within an embedded center, highlighting the distinction between sensor integration and sensor fusion.

As shown in Fig. 7(c), the Luo-Kay model performs data fusion at four different levels: signal, pixel, feature, and symbolic levels. In this architecture, data collected at the sensor level is transmitted to the fusion center, where undergoes layered and sequential fusion processing. As the data is combined in various ways within the fusion center, the information represented by the data increases from the raw data level to the decision level.

2.2.4. Pau model

The Pau model is a data fusion model based on behavioral knowledge, who is a typical hierarchical architecture (see Fig. 7(d) for the specific structure). In the Pau model, feature vectors are first extracted from raw data, and these vectors are then aligned and associated with defined attributes. Data information is combined, analyzed, and clustered in the sensor characteristic fusion and data analysis levels. The final decision stage consists of a set of behavioral rules, which can be extracted through explicit combination outputs.

Additionally, the Pau model is often regarded as a hierarchical approach consisting of three display levels. At the lowest level, each sensor has a vector space that includes coordinate dimensions and measurement parameters. The next level extracts relevant features from the vectors and associates labels with them. The third level associates feature vectors with events and defines the environment model and fusion strategy.

2.3. Theory and methodologies of information fusion

2.3.1. Classical fusion theory

The classic theory of information fusion is built mainly based on mathematical methods of statistical reasoning and estimation, and is commonly used for the fusion processing of incomplete data (i.e., inconsistent data types, low data credibility, incomplete data information, etc.). Table 2 lists the frequently used symbols in this subsection.

(1) Fusion based on probabilistic-modeling
Probabilistic-modeling-based fusion methods were among the earliest information fusion theories applied and are currently the most widely used standard approach in data fusion applications. \(^{53}\) Probability-based data fusion methods typically rely on Bayesian rules to combine prior information and observational information. They use probabilities to describe the information observed and the necessary process information, and employ specific rules to combine them and obtain the final decision information and comprehensive description.

In practice, although probabilistic data fusion methods often rely on Bayesian rules to combine prior information and observational information, fusion methods based on probabilistic modeling are also commonly referred to as “Bayesian fusion”. However, there is no unique way to combine probabilistic information. Methods such as Kalman and extended Kalman filters, sequential Monte Carlo, or function density estimation can all serve as combination rules for multisource information fusion. \(^{54}\)

As a statistical inference method, Bayesian probability theory represents all types of uncertainties using a unified probability measure. For a single source, the Bayesian formula calculates the probability of a given hypothesis being true based on the combination of the prior probability of the hypothesis and the conditional probability of the event/observation, such as representing mutually exclusive events of being healthy (\(H_1\)) or being sick (\(H_2\)), and corresponding explanatory events (observations) denoted as evidence \(E\) (which can represent observations like “environment pollution,” “healthy diet,” or “regular sleep patterns”). Mathematically, Bayesian inference is represented as

\[
P(\text{\(H_i\)} \mid \text{\(E\)}) = \frac{P(\text{\(E\)} \mid \text{\(H_i\)}) \cdot P(\text{\(H_i\)})}{\sum_{j=1}^{2} P(\text{\(E\)} \mid \text{\(H_j\)}) \cdot P(\text{\(H_j\)})} = \frac{P(\text{\(E\)} \mid \text{\(H_i\)}) \cdot P(\text{\(H_i\)})}{P(\text{\(E\)})}
\]

where, \(P(\text{\(H_i\)} \mid \text{\(E\)})\) is the posterior probability of hypothesis \(H_i\) being true given the evidence \(E\) (e.g., the probability of being sick after observing environmental pollution), \(P(\text{\(H_i\)})\) is the prior probability of hypothesis \(H_i\) being true, and \(P(\text{\(E\)})\) is the conditional probability of observing \(E\).

When multiple items of evidence \(E_i\), where \(i \geq 2\), are present, the Bayesian estimator combines these multiple pieces of evidence in a recursive manner to update the probability distribution/density of the system’s state or decision outcome. \(^{55}\) For example, considering whether an email is spam (\(H_1\)) or legitimate (\(H_2\)), the email contains \(n\) observed words/evidence (such as “I am someone,” “transfer money,” “seven-day trip,” “congratulations, you won,” etc.), with probabilities of occurrence in spam emails denoted as \(p(E_1), p(E_2), \ldots, p(E_n)\). The probability of the email being classified as spam can be calculated by combining these multiple pieces of evidence \(E_1, E_2, \ldots, E_n\) using

\[
P(\text{\(H_1\)} \mid E_1, E_2, \ldots, E_n) = \frac{P(H_1) \cdot \prod_{i=1}^{n} P(E_i \mid H_1)}{P(E_1, E_2, \ldots, E_n)}
\]

Nonetheless, both the prior distribution and the normalization term involve integrals that are often difficult to evaluate or observe statistics directly, thereby limiting the applicability of probability-based fusion methods to some extent. As a special type of Bayesian filter, the Kalman filter overcomes this limitation by leveraging the advantage of having an exact analytical solution, which replaces the integrals with the analytical solutions of Bayesian estimates, enabling simpler computation of the fusion results. Thereupon, the Kalman filter has become the most popular probability fusion method due to the simplicity and ease of implementation. Ally in the ointment, like other least squares estimators, the Kalman filter is sensitive to data disrupted by outliers, which makes it unsuitable for applications where the error characteristics are difficult to parameterize. Accordingly, when dealing with data fusion problems in nonlinear systems, further extensions of the Kalman filter are usually based on approximation techniques.

For example, the Extended Kalman Filter (EKF) \(^{56}\) and the Unscented Kalman Filter (UKF) \(^{57}\) are extended models that address data processing problems in nonlinear systems by applying first- and second-order approximations to the original Kalman filter. However, since these methods only approximate the EKF within the first- or second-order range, they can only handle nonlinear problems within limited scope, and thus still face challenges in most nonlinear systems. To address this issue, Stone et al. \(^{58}\) proposed a grid-based method as an alternative approach to approximate nonlinear probability density functions. However, this method suffers from a rapid increase in computational complexity as the data dimensionality increases.

To address the “curse of dimensionality”, the Markov Chain Monte Carlo algorithm emerged. The fundamental idea behind the MCMC algorithm is to alleviate the burden of approximating high-dimensional densities by using a Markov chain to evolve samples, rather than simply performing independent random computations at each step. In MCMC, the Markov chain is a sequence of random samples generated based on a transition probability (kernel) function with Markov properties. This means that the transition probabilities between different sample values in the state space depend only on the current state of the random sample. Therefore, for carefully designed Markov chains, the chain can converge to a specific stationary density, as far as the extracted samples are concerned. \(^{59}\)

(2) Belief functions theory

Belief functions theory traces its origins back to Dempster’s work \(^{60}\) on the reliability of source states in MSIF, aiming to comprehend and refine Fisher’s probabilistic reasoning methods. This theory was later formalized mathematically by Shafer and us as a general theory of evidence-based reasoning. \(^{61}\) encompassing two main components: evidence reasoning \(^{62}\) (Dempster-Shafer Theory, DST) and the Dezert-Smarandache Theory \(^{63}\) (DSmT). The belief functions theory introduces the concept of assigning beliefs and plausibilities to possible measurement hypotheses and the combination rules required to fuse them, which deals with data that involves uncertainty and imprecision and can be considered as an extension of Bayesian theory for handling probability mass functions. In 1981, Garvey et al. \(^{64}\) first applied the theory to information fusion.

The term ‘beliefs’ here encompasses not only common reliability but also engineering-applicable indicators like truthfulness and dependence, which are prevalent in the context of MSIF. \(^{65}\) These indicators interact in this framework and collectively influence the system’s overall performance. Thus, successful information fusion demands a balance of these factors during source information processing to ensure that the system produces accurate and credible results. In particular:
Reliability pertains to the consistency and stability of the information provided by a source. In the context of MSIF, a reliable source is one that delivers consistent information across various times and contexts, resistant to interference or misguidance. Unreliable sources can lead to erroneous judgments or predictions by the fusion system, affecting its accuracy and credibility. If a majority of sources are unreliable, the system might suffer severe consequences, possibly impairing its normal operation.

Truthfulness refers to the degree to which the information provided by a source aligns with reality. False information can result in erroneous judgments by the fusion system, leaving it susceptible to deception when faced with malicious data, leading to unreliable outcomes, and significantly impacting system performance.

Dependence signifies the degree of interconnectedness among sources. In MSIF, sources might exhibit interdependencies, wherein the information from one source is influenced by another. The inter-source dependency determines the information weighting and processing of the system. Overreliance on a specific source while ignoring other sources can lead to information imbalance, affecting the comprehensiveness and accuracy of fusion results, while appropriately considering that the interdependence of sources can enhance the robustness and stability of the system.

In application, unlike probability fusion based on Bayesian inference, the advantage of belief functions theory lies in the ability to provide information at different levels of granularity. In human behavior recognition tasks, identified results can be coarse-grained static and motion states, as well as fine-grained states such as lying, sitting, standing, walking, running, and jumping. Let set \( \Theta = \{ \theta_1, \theta_2, \ldots, \theta_6 \} \) represent the six classes of postures: lying, sitting, standing, walking, running, and jumping. We refer to \( \Theta \) as the recognition framework. Subsequently, using a basic belief assignment function \( m \) (satisfying Eq. (3)), we systematically assign confidence to all potential recognition results.

\[
m(\emptyset) = 0, \quad \sum_{H \subseteq 2^\Theta} m(H) = 1 \tag{3}
\]

In Eq. (3), \( 2^\Theta = \{ \emptyset, \{ \theta_1 \}, \{ \theta_2 \}, \ldots, \{ \theta_6 \}, \{ \theta_1, \theta_2 \}, \{ \theta_1, \theta_3 \}, \ldots, \emptyset \} \) represents the power set of \( \Theta \), which includes all subsets of \( \Theta \) (also known as the power set). \( m(H) \) represents the degree of support of the current evidence \( E \) for proposition \( H \), i.e., the basic belief assignment value. Since there is no state in practice that is similar to “both running and lying”, the corresponding basic belief assignment values \( m \) are set to \( \theta \). Furthermore, when \( m(H) \leq 0 \), \( H \) is referred to as a focal element. In the given example, lying, sitting, standing, walking, running, and jumping are all focal elements.

When fusing multiple sources of information, each input source is considered as evidence with the own independent basic belief assignment function. For instance, when collecting the status information of a human body using \( n \) sensors simultaneously, each sensor output can be treated as an independent evidence source \( E_i, i \in [1, n] \), with a corresponding basic belief assignment function \( m_{E_i}, i \in [1, n] \). The fusion and processing of multiple sources of information can be achieved using the Dempster’s rule, Eq. (4), for combining the evidence. Here, \( K \) represents the conflict coefficient, which is used to measure the conflict size of recognition results given by different evidence sources. The larger the \( K \), the greater the conflict between evidence sources. When \( K=1 \), the rule becomes invalid.

\[
m(H) = \left\{ \begin{array}{ll}
\frac{\sum_{H_i \cap H \neq \emptyset} \prod_{i=1}^{n} m_i(H_i)}{K}, & 1 \leq i \leq n \text{ and } H \neq \emptyset \\
0, & H = \emptyset
\end{array} \right.
\tag{4}
\]

In Eq. (4), \( m(H) \) represents the combined belief assignment for proposition \( H \). The numerator of the equation calculates the belief of \( H \) by considering all possible subsets \( H_i \) of \( H \) and combining the belief assignments from each evidence source. The Dempster’s rule provides a systematic way to fuse and combine multiple sources of evidence, accounting for their respective conflicts and providing a unified belief assignment.

While the Dempster’s rule of DST serves as a valuable approach for representing and handling uncertain information, the rule may yield counterintuitive outcomes when fusing highly conflicting evidence over the system. For example, combining two conflicting sets of evidence might result in an incorrect assignment to the final result, making the system unreliable.

In comparison with the DST, the DSmT exhibits superior capabilities in handling non-exclusive elements, analogous to the D-Numbers Theory (DNT) and order-2 fuzzy sets. The DSmT extends the conventional DST, aiming to address limitations encountered by the DST when dealing with non-exclusive elements, while the pertinent content of order-2 fuzzy sets will be introduced and discussed in the subsequent section.

The core idea of DNT revolves around fusing evidence from various information sources to a set of “basic belief assignment” represented typically using D-functions. These D-functions express uncertainty distributions over different hypotheses (or elements), and the process of fusing these D-functions encompasses three steps: combination, normalization, and uncertainty reassignment.

(1) Combination. Initially, the D-functions from different information sources are combined, and often implemented using combination rules. These rules amalgamate the uncertainties among different information sources, generating a novel D-function.

(2) Normalization. The D-function resulting from the combination might not adhere to the normalization property, i.e., the sum of probabilities is not equal to 1. Normalization ensures that the fused D-function constitutes a valid probability distribution.

(3) Uncertainty reassignment. In the DS theory, different elements are mutually exclusive. However, in certain cases, elements might not be mutually exclusive, necessitating a reassignment of uncertainty to accommodate this scenario.

Benefiting from the uncertainty allocation strategy of D-functions, DNT possesses advantages in handling non-exclusive elements (the DNT is more suited for addressing non-exclusive elements relative to traditional DST, where elements could potentially overlap) and offers enhanced expressive capabilities (DNT introduces novel concepts such as “non-specificity” and “nonspecificity” to better represent diverse aspects of uncertainty). Consequently, DNT holds
a certain advantage in dealing with non-exclusive elements and complex uncertainty situations. However, this advantage needs to be balanced against computational complexity (similar to the DST, DNT may involve intricate computations during information fusion, especially when dealing with a large number of information sources) and data requirements (constructing basic belief assignment functions in DNT necessitates sufficient data; inadequate information sources could lead to inaccurate fusion outcomes). In specific contexts, DNT can prove to be a valuable tool for information fusion and uncertainty reasoning.

In addition to conflicting evidence and the fusion of non-exclusive elements, uncertainty reasoning has remained a focal point of interest in the practical application of credibility function theory within the domain of Multiple Sources of Information Fusion (MSIF). A representative approach is the Complex Evidence Theory (CET), known for its strong interpretability and recent attention. For instance, in Ref. [72], aiming to enhance the uncertainty reasoning performance of expert systems, a generalized correlation coefficient named Complex Evidence Correlation Coefficient (CECC) was proposed for complex quality functions or Complex Basic Assignment (CBBAS) in CET. Building upon this, a complex conflict coefficient was introduced to measure conflicts among CBBAS. In this context, a weighted discounting multi-source information fusion algorithm, CECC-WDMSIF, was devised based on CECC to enhance the performance of CET-based expert systems. In Ref. [73], to better capture uncertainty within knowledge, the expression of CET was framed within the quantum framework of Hilbert space. Consequently, a generalized negation method for quantum basic assignments was proposed. Through the investigation of the negation process, trends in negation iteration were revealed, offering a promising solution for knowledge representation, uncertainty measurement, and quantum information fusion.

Different from fusion based on probabilistic modeling, which must assume a uniform distribution to deal with this unknown problem, although the belief functions theory only allocates probability when there is supporting information (that is, does not assign prior probability to unknown propositions), the reliability function theory allows to express the confidence of uncertain information by assigning the whole mass to the discrimination framework, that is, to meet $m(E = H) = 1$ at any time. As a result, when facing a specific task, the choice of fusion methods based on probabilistic modeling and belief functions theory essentially involves balancing the desire for higher accuracy in data (fusion based on probabilistic modeling) against the need for more flexible fusion formulas (belief functions theory).

(3) Fuzzy set theory

Taking the evaluation of food taste as an example, although precise quantification of ion concentrations can be achieved through instruments like an electronic tongue, the common perception among the general public is often expressed using relative and vague terms such as “bad” “average” or “delicious”. Moreover, the same dish can receive completely different evaluations, such as “bland” “mild” “suitable” “salty” or “rich” when different diners are involved (evaluation variations due to changes in evidence sources). Therefore, for tasks that involve imprecise reasoning, the observation results provided by different sensors for the same target are often fuzzy and uncertain. Hence, when combining multiple information sources with uncertainty, it is necessary to perform information fusion based on fuzzy reasoning principles.

A fuzzy set $F \subseteq X$ is defined by the corresponding gradual membership function $\mu_F(x)$ ($\mu_F(x) \in [0, 1]$) for all $x \in X$, which represents the degree of membership of an element $x$ in $X$ to $F$. The higher the membership degree, the stronger the affiliation of $x$ to $F$, which makes fuzzy data fusion an effective solution for dealing with imprecise data. Specifically, fuzzy data is first fuzzified using gradual membership functions. Then, fuzzy fusion output is generated by combining the fuzzy data using fuzzy rules. Fuzzy fusion rules can be categorized into conjunction and disjunction.

In fuzzy set theory, conjunction and disjunction Eq. (6) are operations used to combine the membership degrees of elements in two fuzzy sets. These concepts are similar to the “AND” and “OR” operations in classical logic, but in fuzzy logic, they deal with degrees of membership (values between 0 and 1), rather than absolute truth or falsehood.

Specifically, let $F_1$ and $F_2$ be two fuzzy sets, with $\mu_{F_1}(x)$ and $\mu_{F_2}(x)$ representing the membership degrees of an element $x$ in $F_1$ and $F_2$, respectively. Then, the conjunction $F_1 \cap F_2$ (or $F_1 \text{AND} F_2$) of $F_1$ and $F_2$ is defined by the membership function Eq. (5), which means that the membership degree of $x$ in the conjunctive set is the minimum of its membership degrees in the two fuzzy sets.

$$\mu_{F_1 \cap F_2}(x) = \min(\mu_{F_1}(x), \mu_{F_2}(x)) \quad (5)$$

Similarly, for two fuzzy sets $F_1$ and $F_2$, the disjunction $A \cup B$ (or $A \text{OR} B$) is defined by the membership function:

$$\mu_{F_1 \cup F_2}(x) = \max(\mu_{F_1}(x), \mu_{F_2}(x)) \quad (6)$$

Where, the membership degree of $x$ in the disjunctive set is the maximum of its membership degrees in the two fuzzy sets.

These definitions are fundamental operations in fuzzy logic for handling fuzzy sets, forming the basis for more complex fuzzy logic structures and fuzzy rule systems. By applying these fuzzy fusion rules, the combination of imprecise data can be effectively achieved, providing a robust solution for information fusion in the presence of uncertainty. While the minimum and maximum operations are the most common for conjunction and disjunction in fuzzy logic, other methods, such as algebraic product or sum, are also used.

Differently, different applications and contexts may choose different methods for handling the conjunction and disjunction of fuzzy sets.

When fusing homogeneous sensor data with equal levels of credibility, joint fuzzy fusion rules are commonly used. However, there are cases where at least one reliable data source is unknown or when highly conflicting data needs to be fused, making it difficult to simultaneously apply conjunction and disjunction rules. To meet the challenge, some researchers have proposed adaptive fuzzy fusion rules as a compromise when joint fuzzy fusion rules cannot be used. For instance, Eq. (7) presents an adaptive fuzzy fusion rule for these two situations.

$$\left\{ \begin{array}{l} \mu_{\text{Adaptive}}(x) = \max\{\mu_1, \mu_2\} \quad \forall x \in X \\ \mu_1 = \frac{\mu_{F_1 \cap F_2}(x)}{h(\mu_{F_1}(x), \mu_{F_2}(x))} \\ \mu_2 = \min\{1 - h(\mu_{F_1}(x), \mu_{F_2}(x)), \mu_{F_1 \cup F_2}(x)\} \end{array} \right\} \quad (7)$$

Here, $h(\mu_{F_1}(x), \mu_{F_2}(x))$ is used to measure the degree of conflict between the gradual membership functions $\mu_{F_1}(x)$ and $\mu_{F_2}(x)$. $\mu_1$ and $\mu_2$ represent the desired conjunction and disjunction fuzzy fusion rules, respectively. The adaptive fuzzy fusion rule in Eq. (7) allows for a flexible combination of fuzzy sets in situations where traditional conjunction and disjunction rules cannot be simultaneously applied. By considering the conflict between membership functions, the adaptive rule provides a viable approach for information fusion in scenarios with unreliable sources or highly conflicting data.
In the realm of MSIF tasks involving complex uncertain information, the order-2 fuzzy set theory, an extension of the traditional fuzzy set theory, emerges as a powerful framework for describing relationships among different membership functions with enhanced precision, which proves particularly effective in handling uncertainty with a finer granularity. Within the framework of order-2 fuzzy set theory, membership functions themselves are treated as fuzzy sets, thereby categorizing first-order membership functions as a subset of order-2 membership functions. Mathematically, the order-2 fuzzy set theory can be defined as follows:

(A) First-Order Membership Function: Let the first-order membership function of an element \( x \) belonging to fuzzy set \( A \) be denoted as \( \mu_A(x) \), where \( x \) lies within a specific interval or set, and \( \mu_A(x) \) takes values within the range of \([0, 1]\).

(B) Second-Order Membership Function: For a first-order membership function \( \mu_A(x) \), its distribution across the entire membership space can be represented by another membership function \( \mu_B(\mu_A(x)) \) where \( B \) denotes the order-2 fuzzy set that represents the distribution of membership.

For instance, if \( \mu_A(x) \) signifies the extent to which an element \( x \) belongs to a certain temperature range, \( \mu_B(\mu_A(x)) \) can indicate the degree to which \( x \)'s temperature membership is within a specific range. Thus, the order-2 fuzzy set \( \mu_B \) can be defined as

\[
\mu_B(\mu_A(x)) = y
\]

where \( y \) represents the degree of the temperature membership distribution of \( x \) within set \( B \). This approach offers a more precise representation of membership distribution relationships, enabling a more accurate representation of uncertainty information.

Comparatively, the order-2 fuzzy set theory surpasses traditional fuzzy set theory in uncertain environment information fusion tasks, owing to its advantages in more accurate uncertainty modeling (by better describing relationships among different membership functions, allowing for precise uncertainty representation in complex scenarios) and improved information fusion (leveraging second-order membership functions for more effective integration of information from diverse sources, leading to comprehensive and precise outcomes). However, this also demands greater computational efforts and model design complexities. The incorporation of order-2 fuzzy set theory introduces increased computational complexity due to the involvement of higher-dimensional membership functions. Moreover, designing order-2 fuzzy set models tailored to specific problems necessitates more domain expertise and experience. Consequently, practical application warrants a thorough consideration of the theory’s pros and cons, weighed against the complexity of the problem and the desired precision.

To sum up, although fuzzy set theory, whether order-2 or traditional, similar to fusion based on probabilistic modeling, requires prior knowledge of membership functions for different fuzzy sets, but offers a powerful framework for modeling the fuzzy membership degrees of targets in situations where the definition of objects is unclear. Unlike fusion based on probabilistic modeling and evidence theory, which are well suitable for modeling uncertainty within well-defined object categories, fuzzy set theory excels at capturing uncertainty in contexts where object definitions are ambiguous. Moreover, due to the ability to represent and fuse linguistic fuzzy data generated by human experts, fuzzy set theory proves particularly useful. While probability theory and evidence theory rely on precise numerical values, fuzzy set theory accommodates the inherent imprecision in human expert knowledge, allowing for the representation and fusion of linguistic terms such as “very high” “medium” or “low”, which makes fuzzy set theory a valuable tool for handling and integrating subjective and imprecise information in various domains. In a word, fuzzy set theory is a robust method for modeling and fusing uncertain and fuzzy data, which is very suitable for dealing with the challenges brought by the lack of accurate object definitions and human expertise.

(4) Rough set theory

Rough set theory is a mathematical tool for handling imprecise, uncertain, and incomplete data, which allows for the analysis of data without the need for additional prior knowledge, unlike fusion based on probabilistic modeling which requires additional prior knowledge for data analysis, and which enables the treatment of data granularity while ignoring the uncertainty of different granularities. Rough set theory provides a method to approximate the target set \( T \) within a given framework \( F_B \) for a set \( B \subseteq A \), where \( F_B \) describes a specific set of attributes for the objects. The approximation is represented as the tuple \( < B_1(T), B^*(T) > \), where \( B_1(T) \) and \( B^*(T) \) denote the lower and upper approximations of the set \( T \) within the framework \( F_B \) (i.e., \( B_1(T) \subseteq T \subseteq B^*(T) \)). The definitions of the lower and upper approximations are given by

\[
\begin{align*}
B_1(T) &= \{ O \mid \{ O \}_{F_B} \subseteq T \} \\
B^*(T) &= \{ O \mid \{ O \}_{F_B} \neq T \neq \emptyset \}
\end{align*}
\]

The lower approximation \( B_1(T) \) can be interpreted as a conservative approximation, including only objects that definitely belong to \( T \), while the upper approximation \( B^*(T) \) includes all objects that possibly belong to \( T \). Based on this approximation, the boundary region of \( T \) is defined as \( BN_B(T) = B^*(T) - B_1(T) \), which consists of objects that cannot be classified as either belonging or not belonging to \( T \). If \( BN_B(T) \neq \emptyset \), then the set \( T \) is considered rough.

In the information fusion model, the set \( T \) is considered an imprecise set representing the system state, and rough set theory allows for the approximation of the system’s possible states based on the granularity of the input data, denoted as \( F_B \). Once approximated as a rough set, data segments can be fused using classical set-theoretic conjunction or disjunction operators, corresponding to intersection and union, respectively. To achieve effective fusion, the granularity of the data should neither be too fine-grained nor too coarse. In the case of fine-grained data, where \( \{ O \}_{F_B} \) is a singleton, rough set theory is simplified to classical set theory. On the other hand, when dealing with very coarse data granularity, where \( \{ O \}_{F_B} \) is a large subset, the lower approximation of the data may become empty, resulting in complete “ignorance”. Overall, the primary advantage of rough set theory, compared to other fusion theories, is that it does not require any prior or additional information, such as probability distributions or membership functions of the data. Additionally, rough set theory allows for fusion based solely on the internal structure (granularity) approximation of imprecise data.

In practical applications, such as the purchase of a car, customers consider multiple features (e.g., brand, price, color, emission, etc.), each of which represents input information from different sources. In this scenario, each feature can be transformed into an attribute, and a set can be defined for each attribute feature, mapping to corresponding attribute values. For example, car brand can be defined as three sets: “low-end brand” “mid-range brand” and “high-end brand”. It’s worth noting that the granularity of different attributes may vary, and rough set theory allows for information fusion across different granularities. Using rough set theory, multiple attributes from different sources can be combined to determine the purchasing target, which meets requirements such as budget, brand, color, etc.

(5) Comparative analysis of classical fusion theories
As the four most classical fusion theories in MSIF, probabilistic modeling, belief functions theory, fuzzy set theory, and rough set theory have their advantages in different application scenarios. Based on the existing literature, we have summarized their respective strengths and weaknesses in Table 3.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessary</td>
<td>Massive computing power</td>
<td>✓</td>
<td>✓</td>
<td>Ø</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Large datasets</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Priori knowledge</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Whether with</td>
<td>Explainability</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Information with noise</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Multi-granularity information</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Ability to process</td>
<td>Fuzzy information</td>
<td>×</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Conflicting information</td>
<td>✓</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Uncertain information</td>
<td>√</td>
<td>✓</td>
<td>×</td>
<td>√</td>
</tr>
</tbody>
</table>

Note: In the table, "✓" means "yes", "×" represents "no", and "Ø" indicates that there is controversy or there is currently no clear definition of relevant reports.

(A) Computational demand

Fusion based on probabilistic modeling: Typically involves complex probability computations, especially in high-dimensional spaces, resulting in significant computational overhead.
Belief functions theory: Entails high computational complexity, particularly with large-scale evidence sets.
Fuzzy set theory: Generally features moderate computational complexity, with complexity potentially increasing for highly fuzzy information.
Rough set theory: Offers relatively simpler computations, making it applicable to larger datasets.

(B) Training data requirements

Fusion based on probabilistic modeling: Demands substantial training data to estimate probability distributions accurately.
Belief functions theory: Requires sufficient evidence to construct belief functions.
Fuzzy set theory: Demands less training data, relying primarily on the specification of fuzzy membership functions.
Rough set theory: Emphasizes data granularity rather than extensive training data.

(C) Reliance on prior knowledge

Fusion based on probabilistic modeling: Relatively low dependence on prior knowledge, primarily driven by data-driven learning.
Belief functions theory: The prior knowledge required to flexibly combine subjective and objective information in certain specific situations.
Fuzzy set theory: Limited reliance on prior knowledge, primarily facilitated by fuzzy membership functions.
Rough set theory: Minimal dependence on prior knowledge, emphasizing rough relationships among data.

(D) Explainability

Fusion based on probabilistic modeling: Generally offers good explainability, as fusion results can be explained through probability-based reasoning.
Belief functions theory: Demonstrates robust explainability only in terms of expert-provided trust levels and uncertainty assessments.
Fuzzy set theory: Effectively manages fuzzy and imprecise information, though explainability might be somewhat intricate.
Rough set theory: Provides explainability through rough set characteristics, aiding in understanding rough relationships.

(E) Handling noisy source information

Fusion based on probabilistic modeling: Possesses a degree of robustness against noise, proficiently managing noise through probability distributions.
Belief functions theory: Adequate in handling noise to a certain extent, but efficacy might be challenged when dealing with significant conflicts.
Fuzzy set theory: Tolerates noise to a certain extent, though impact may be observed in highly noisy scenarios.
Rough set theory: Has the ability to manage low noise to a certain extent, and the effectiveness will decrease as the noise intensity increases.

(F) Multi-granularity information fusion

Fusion based on probabilistic modeling: Effectively manages multi-granular information, e.g., through Gaussian mixture models.
Belief functions theory: Capable of handling multi-granular information by combining belief functions.
Fuzzy set theory: Multi-granularity information can be fused by defining membership functions with different granularities. However, membership functions with different granularities are difficult to define, and in some cases, this theory may not capture the complex relationships of information well. Therefore, there is some academic controversy over its applicability to multi-granularity information fusion.
Rough set theory: More suitable for coarse-grained information partitioning.

(G) Incorporating fuzzy information

Fusion based on probabilistic modeling: The inapplicability to tasks involving fuzzy information stems from the risk of losing crucial information by naively mapping fuzzy data into probability spaces. Traditional probabilities are characterized via density functions or distributions, whereas fuzzy information may necessitate alternative representations of its inherent fuzziness. Unlike probabilities confined within defined ranges, membership degrees in fuzzy sets can be distributed across multiple sets. Fusing fuzzy information entails non-random uncertainty, where indistinct relationships are not solely induced by stochastic events. This contrasts with the probabilistic concept of randomness.
Belief functions theory: Lack of sufficient literature to discuss this characteristic.
Fuzzy set theory: Inherently suitable for fusing fuzzy information.
Rough set theory: Not adept at handling fuzzy information.

(H) Conflict resolution in source information

Fusion based on probabilistic modeling: Only when the reliability of different sources is known, can information conflicts be addressed through source weighting.
Belief functions theory: Proficiently addresses conflicting information by managing inconsistent evidence (but with the risk of failure in the event of complete conflict between sources).
Fuzzy set theory: Possess the ability to handle conflicting information only when the membership function design is reasonable.
Rough set theory: Handles rough conflicts among sources, but
generally consists of four steps: input data, feature extraction, modeling
training, and result output. The basic model shown in Fig. 9 represents a traditional machine learning model or deep learning model (such as support vector machines or convolutional neural networks) that is capable of achieving end-to-end input-output data mapping, which is used to extract the features or decision information to be fused. The information fusion component represents a novel information integration model constructed from one or multiple basic models.

Compared to classical fusion theories, AI can provide more accurate descriptions and precise decision results for imperfect data detection without introducing prior knowledge, which is achieved through data preprocessing and segmentation, feature extraction and dimensionality reduction, as well as learning optimization of classification/clustering models. AI based information fusion methods have gained significant importance in practical engineering applications. While AI has significantly enriched the technical approaches for MSIF, which has also brought forth new challenges and opportunities. For instance, while MSIF focuses on providing objective descriptions or comprehensive evaluation methods for scenarios and impacts, AI primarily focuses on big data analysis for classification assessment, which raises questions regarding the integration of the two, algorithm selection, and how fusion should be defined, which need to be analyzed in specific applications.

This section begins with a brief description of the basic models for implementing MSIF tasks using deep learning, including traditional machine learning and the four widely-used deep learning methods, and provides an overview of multi-view learning, which has achieved fruitful results in cross-modal/multi-modal information fusion. Furthermore, we introduce the transfer learning, which is widely applied in the field of deep learning. Finally, this section concludes with a brief summary and analysis of the theoretical feasibility, fusion strategies, and applicable domains of deep learning for MSIF.

(1) Traditional machine learning

Distinguished from deep learning which combines feature extraction and modeling training into a single process, traditional machine learning typically performs these two steps separately as needed (see Fig. 9). Traditional machine learning, as a technique with powerful data computation and classification capabilities, is expected to improve the overall performance of data fusion algorithms and may even have higher data processing efficiency in certain application scenarios, despite being more cumbersome in form compared to deep learning. Currently, traditional machine learning algorithms used for MSIF primarily include support vector machines (SVM), k-nearest neighbors, Naive Bayes, random forests, hidden Markov model, etc.

In specific MSIF applications: Banerjee et al. employed support vector machines to construct a data-level MSIF model for hybrid fault detection, achieving detection and identification of three states: healthy, degraded, and failed, in the detection system; Razavi-Far et al. proposed a similarity learning information fusion approach based on the k-nearest neighbors algorithm for missing data imputation tasks and demonstrated the effectiveness of the method on 11 public datasets; Gao et al. addressed the task of multi-view human behavior recognition and developed a category-level dictionary learning model with feature-level fusion for adaptive fusion of different view data, significantly improving the recognition accuracy of multi-view action; Li et al. applied naive Bayes to fuse multimodal information to describe a robot’s dance pose and estimate the aesthetic appeal of the robot’s dance pose; to tackle the challenge of producing the same product in different batches under the same working conditions in intermittent industrial processes, which often involve complex and continuous
Multi-source information fusion: Progress and future

13

Information fusion based on AI model

Traditional machine learning

Support vector machines e.g. Ref.89
K-nearest neighbors e.g. Ref.90
Dictionary learning e.g. Ref.91
Naive Bayes e.g. Ref.92
Others e.g. Hidden Markov model (Ref.93)

Deep learning

Convolutional neural network e.g. Refs.101,102
Recurrent neural network
Traditional RNN e.g. Ref.105,106
LSTM e.g. Ref.108
GRU e.g. Ref.109
Others e.g. Ref.110
Transformer e.g. Refs.128-130

Transfer learning e.g. Refs.161,162

Others e.g. Ref.110

Multiview learning e.g. Refs.132-141
Transfer learning e.g. Refs.161,162

Feature engineering (extraction) and modeling training can be conducted separately or synchronously

Other approaches

Knowledge representation and reasoning e.g. Ref.81
Intelligent computing e.g. Ref.82
Expert systems e.g. Ref.83
Rule engines e.g. Ref.84
Genetic algorithms e.g. Ref.85
etc.

Feature engineering and modeling training can be conducted separately or synchronously

Synchronize feature extraction during modeling training

Fig. 8 Illustration of MSIF approaches based on AI.

Physicochemical reactions, Sun et al.98 utilized a stacked multi-output random forest algorithm to establish the predictive relationship between key process variables at specific stages and multiple quality indicators, proposing a multi-stage information fusion strategy for industrial batch data-driven quality prediction; Yang et al.99 employed a first-order hidden Markov model to compute the dynamics of Bayesian networks at different time slices, fusing context information of physiological features (e.g., ECG and EEG) to achieve fatigue driving detection.

These MSIF methods based on traditional machine learning are all constructed based on simple mathematical models, so their results have strong explainability and can provide intuitive understanding of decision-making. In the case of relatively small data volume, it has the advantages of easy implementation and good generalization performance. However, they also have the common problem of traditional machine learning algorithms, which is that the demand for feature engineering is relatively high, which may become cumbersome and time-consuming when facing a large amount of different types of multi-source data, and may not be able to effectively model complex nonlinear relationships. Therefore, in some specific application scenarios (e.g., large-scale data or complex relationship modeling), it is necessary to study the MSIF of artificial intelligence algorithms based on non-traditional machine learning such as deep learning.

(2) Convolutional neural network

The Convolutional Neural Network (CNN) is a deep learning architecture inspired by the biological visual perception mechanism.100 Since 2012, CNN has achieved breakthrough results in various fields, ranging from image processing to speech recognition and natural language processing. On the one hand, CNN effec-
tively reduces the number of parameters in deep neural networks, enabling the use of larger model capacities to tackle more complex tasks.\textsuperscript{104} On the other hand, in MSIF tasks involving time series decision-making, CNN exhibits characteristics such as local dependency and scale invariance, which make CNN superior to other machine learning methods and highly favored by researchers.\textsuperscript{102}

In the $l$-th convolutional layer of the CNN model, let $X(1-l) \in \mathbb{R}^{m \times n \times C_{in}}$ denote the input features, $K(l) \in \mathbb{R}^{m \times n \times C_{in}}$ represent $C_{out}$ convolutional kernels of size $n \times n$, and $B(l) \in \mathbb{R}^{C_{out}}$ denote the bias elements. The function $f$ represents the activation function. The output features of this convolutional layer, denoted as $X(l) \in \mathbb{R}^{m \times n \times m \times n \times C_{out}} (r \text{ denotes the stride of the convolution operation})$, can be computed for each element $(l)_{u,v,w} (1 \leq u,v \leq m, 1 \leq w \leq C_{out})$ using

$$\begin{align*}
\chi(i,j) &= 1, 0 \leq i,j \leq n \\
\chi(i,j) &= 0, \text{ others} \\
(l)_{u,v,w} &= f((l)_{u,v,w})
\end{align*}$$

In recent years, CNN has achieved remarkable results in numerous MSIF tasks. For instance, CNN-based fault diagnosis models, which allow participants from different industries to collaborate on training a global fault diagnosis model without sharing their local data via federated learning, have been developed, i.e., the participants’ local CNNs are aggregated through cloud servers to update the global model, enabling collaborative training of the global fault diagnosis model in a homogeneous MSIF setting.\textsuperscript{103}

Benefiting from the powerful feature extraction and encoding capabilities, CNN is well suited not only to homogeneous information fusion tasks but also to heterogeneous information fusion tasks. As a result, CNN has made significant contributions in the field of multimodal information fusion applications. For example, Liu et al.\textsuperscript{104} employed CNN to perform heterogeneous information fusion by combining ultrasound images and RF signals, achieving improved classification of thyroid nodules.

(3) Recurrent neural network

Recurrent Neural Networks (RNN) are well suited to handling time-series data, e.g., audio and text.\textsuperscript{106} RNNs have connections between hidden layer units, allowing information to pass from one neuron to the next, enabling the extraction of temporal relationships.\textsuperscript{107} However, RNNs suffer from the vanishing gradient problem when dealing with long input sequences, which hinders their ability to capture long-range dependencies. To address this issue, several advancements have been made to improve RNNs, with the most effective approach being the introduction of gate mechanisms, such as Long Short-Term Memory (LSTM) networks.\textsuperscript{108} Gated Recurrent Unit (GRU) neural networks,\textsuperscript{109} and convolutional recurrent neural network.\textsuperscript{110}

LSTM\textsuperscript{111} is a special type of RNN that addresses the problem of gradient vanishing or exploding. A common issue in deep networks is the “vanishing gradient” problem, where the gradient becomes smaller with each layer, eventually becoming too small to affect the deepest layers. The memory cells of LSTM allow for continuous gradient flow, preserving the error value, thereby eliminating the vanishing gradient problem and enabling the learning of long-term dependencies from sequences spanning hundreds of time steps. Zhang et al.\textsuperscript{108} focused on predicting the stock prices of BGI Genomics and applied LSTM with an attention mechanism to identify long-term temporal dependencies and adaptively highlight key features. The experimental results demonstrated that LSTM with attention mechanism significantly improves prediction performance and enhances analysis of time-frequency features through the fusion of heterogeneous information sources, such as online news.

In addition to utilizing LSTM independently for feature handling and information fusion, there has been extensive application of combining CNN and LSTM for feature extraction. Kavi et al.\textsuperscript{112} discussed the integration of LSTM and CNN models for multi-view data fusion. Chen et al.\textsuperscript{110} designed a multi-source homogeneous information fusion model using CNN and bidirectional LSTM with two parallel hybrid branches. This model simultaneously focuses on beat-based information in Electrocardiogram (ECG) rhythms and segment-based information in adjacent beat segments for automatic arrhythmia detection. Nii et al.\textsuperscript{113} constructed a hybrid deep neural network architecture based on LSTM and CNN to fuse multiple heterogeneous data sources, including public sentiment and economic factors, for predicting the stock prices of the Ghana Stock Exchange.

GRU\textsuperscript{114} was designed with the same intention as LSTM, to address the issues of vanishing or exploding gradients during the processing of long sequential data by RNN models. Thereupon, in most practical tasks, GRU and LSTM exhibit similar performance and fitting capabilities, but GRU offers the advantage of being computationally more efficient than LSTM, making it widely used GRU certain scenarios.\textsuperscript{115} Zhao et al.\textsuperscript{116} input two travel time values (dedicated short-range communication data and remote traffic microwave sensor data) into a GRU model to obtain travel time predictions after MSIF, thereby achieving road traffic condition prediction.

Moreover, similar to LSTM, there have been recent efforts to combine GRU with CNN for decision fusion tasks involving multiple information sources. For instance, Kong et al.\textsuperscript{117} fused spatial-temporal features using a structure combining CNN and GRU to monitor wind turbine conditions. Guo et al.\textsuperscript{118} employed multi-scale CNN to extract spectral features and fused them with time-series features extracted by GRU for diagnosing Hepatitis B virus infections. Zhang et al.\textsuperscript{119} integrated CNN and GRU models to improve the accuracy of environmental sound classification by fusing features extracted from logarithmic mel-spectrograms, mel-frequency cepstral coefficients, and raw waveforms using CNN-GRU.

(4) Transformer

The Transformer\textsuperscript{120} is a deep learning architecture that integrates the attention mechanism into a feed-forward neural network, which consists of self-attention and feed-forward modules in a single-layer structure, with the feed-forward module being approximated as a residual module with normalization operations. Given an input feature $X$, the self-attention mechanism is modeled as follows: the self-attention mechanism linearly/non-linearly maps the input to obtain query ($Q$), key ($K$), and value ($V$) features. Then, the output feature of self-attention is computed based on

$$\text{Attention}(Q,K,V) = \text{SoftMax} \left( \frac{QK^T}{\sqrt{d_k}} \right) \cdot V$$

$$\text{SoftMax}(K) = \frac{\exp(k_i)}{\sum_{k_i} \exp(k_i)} \quad K = [k_1,k_2,\ldots,k_{d_k}]$$

Usually, $K, Q, V$ are tensors with the same size, i.e. $K, Q, V \in \mathbb{R}^{B \times L \times d_k}$, where $B$ represents the number of samples in a batch, $L$ is the length of the processed tensor, and $d_k$ is the dimension of Transformer, and $\sqrt{d_k}$ is a scaling factor usually equal to the square root of the dimension of $K$.

Like LSTM, Transformer is capable of capturing long-range dependencies in sequences, but with a larger context window to handle even longer sequences, and exhibits easier parallelization.
and high scalability.\textsuperscript{121} Therefore, the initial motivation behind the Transformer architecture was to address the challenge of analyzing long sequences that traditional architectures such as LSTM struggled with, and has shown remarkable performance improvements in various text processing tasks based on large-scale corpora.\textsuperscript{122}

Subsequently, the emergence of the Vision Transformer (ViT)\textsuperscript{123} successfully extended the application domain of Transformer from one-dimensional sequences, such as text and audio, to two-dimensional image and three-dimensional video data. With outstanding achievements in tasks such as image classification,\textsuperscript{124} object detection,\textsuperscript{125} and image generation,\textsuperscript{126} ViT has become a hot research direction.\textsuperscript{127}

The Transformer has demonstrated outstanding performance in various tasks involving different types of data, including video, text, images, which showcases the powerful analysis and integration capabilities of Transformer for multimodal data. As a result, in many heterogeneous data fusion tasks involving multimodal information, Transformer is expected to become a highly competitive tool for information processing and fusion.

In the task of affective behavior analysis, Zhang et al.\textsuperscript{128} proposed a fusion approach based on Transformer to handle multimodal information. They encoded static visual features from current frame images and extracted three types of multimodal features from image, audio, and text sequences. By fusing static visual features with dynamic multimodal features, their framework achieved the state-of-the-art performance in the EXPR and AU tracks of the ABAW3 Competition’s official test. For sentiment analysis tasks, Wang et al.\textsuperscript{129} introduced a novel fusion method called TransModality based on Transformer. They extracted features from text, visual, and acoustic modalities and performed feature-level multimodal sentiment interaction fusion for multimodal sentiment analysis. Their approach achieved state-of-the-art performance on the CMUMOSI, MELD, and IEMOCAP datasets at that time. In the field of image fusion, the MATR model\textsuperscript{130}, a Transformer-based scale-adaptive multimodal image fusion model, leverages multimodal information from different scales to perform image fusion. MATR model benefits from complementary fusion of global contextual information and modeling of long-range dependencies, surpassing state-of-the-art models on several mainstream medical image databases.

(5) Multiview learning

Multiview data refers to data captured from different modalities, sources, spaces, or other forms, which share similar high-level semantics. Specifically, multiview data encompasses various forms, e.g. text, video, audio, etc., used to describe the same object, different languages used to report the same event, the same human behavior captured from different camera angles, and the visual information and user tags contained in the same social image.\textsuperscript{131}

Although deep learning models such as CNNs, RNNs, and Transformers have achieved remarkable results in the MSIF task, their models often focus only on single-view information. In real-world MSIF tasks, a single view of information cannot often provide a comprehensive and accurate explanation, and thus limitation leaves ample room for improvement when directly applying these models to handle MSIF tasks. To address this issue, researchers have introduced the concept of MultiView Learning (MVL), which aims to integrate information from multiple views to provide more comprehensive and accurate descriptions and explanations. This approach enables more efficient fusion of multiple sources of homogeneous/heterogeneous information.

Traditional machine learning and deep learning algorithms often concatenate multiview data into single-view data to accommodate the learning setting. However, such concatenation lacks practical physical meaning and can lead to overfitting, especially in scenarios with small training samples, thereupon then it is challenging to meet the fusion and analysis requirements of multiview data.\textsuperscript{132} Simultaneously, in the past few decades, multiview data has become one of the main types of data on the Internet, with the quantity exploding in various domains such as video surveillance,\textsuperscript{133} entertainment media,\textsuperscript{134} social networks,\textsuperscript{135} and medical diagnostics.\textsuperscript{136} Against this, the demand for multimodal and cross-modal information processing and fusion of multiview data has driven the rapid development of multiview learning. MVL aims to achieve joint training by combining different features or data sources based on a common feature space.\textsuperscript{137} Currently, the mainstream MVL method is to map multiview data to a shared feature space that maximizes the mutual consistency between multiple views, as demonstrated in Ref.\textsuperscript{138}

In the past few decades, MVL has shown significant progress in both traditional machine learning and deep learning fields, inspiring numerous promising algorithms, including co-training,\textsuperscript{139} multi-kernel learning,\textsuperscript{140} and subspace learning.\textsuperscript{141}

Co-training, one of the earliest MVL solutions, is a type of “divergence-based” semi-supervised learning method. Co-training first trains a classifier on each view using labeled data. Then, these classifiers are used to generate pseudo-labels for unlabeled samples, which are provided to other classifiers as additional training samples. This process is iteratively repeated, generating new samples alternately, to facilitate the “mutual learning” among multiple classifiers until the classifier weights become relatively stable or reach a preset number of iterations. Co-training relies on three main assumptions: (A) sufficiency, which states that each view is sufficient for individual classification; (B) compatibility, which assumes that co-occurring features with the same label prediction in two views have a high probability; (C) conditional independence, which assumes that given the labels, the views are conditionally independent. Among these assumptions, conditional independence is crucial but often challenging to perfectly satisfy in practice. Therefore, in recent years, many researchers have been dedicated to designing alternative approaches for scenarios with low conditional independence.\textsuperscript{142}

Multi-kernel learning, due to the natural correspondence to different views by constructing multiple kernel functions in high-dimensional space, and the advantage of improving learning performance through linear or nonlinear combinations of kernels, has been widely applied to handle multi-view data processing problems. Based on the construction methods and characteristics of multi-kernel functions, multi-kernel learning can be roughly divided into three categories: composite kernel methods, multiscale kernel methods, and infinite kernel methods. Composite kernel methods involve combining multiple kernel functions with different characteristics to obtain a composite kernel function that is compatible with the characteristics of each sub-kernel function. Multiscale kernel methods, on the other hand, aim to address the challenge of inflexibility in decision-making capability and the lack of scale selection ability in traditional composite kernel methods. They propose a kernel function with multiscale expressive power. Intuitively, this approach involves sequentially learning and fusing different-scale kernels to construct a composite decision function. Considering that composite kernel methods and multiscale kernel methods are both based on combining a finite set of kernel functions (usually through linear combinations), they may face difficulties in representing certain large-scale decision problems uniquely. To address this issue, infinite kernel methods have emerged. These methods optimize the measurement of a minimum regularization function representation through a set of multiple kernel functions, achieving a semi-optimized solution for complex learning problems.
involving general kernels.\textsuperscript{143}

Subspace learning aims to obtain a shared latent subspace among multiple views by assuming that the input views are generated from this underlying subspace. Once a specific subspace is identified, subsequent tasks such as classification\textsuperscript{144} and clustering\textsuperscript{144} can be directly performed. Subspace learning effectively mitigates the “curse of dimensionality” since the dimensionality of the latent subspace is lower than that of any input view.

In recent years, the rapid development of deep learning has not only revolutionized numerous technologies in academia and industry but also diversified the advancements in multi-view learning techniques. On one hand, MVL has benefited from the outstanding feature extraction and recognition capabilities of models such as CNN,\textsuperscript{145} RNN,\textsuperscript{146} and Transformer,\textsuperscript{147} leading to breakthrough progress and outstanding achievements in various tasks. Moreover, just as the development trend of deep learning has transitioned from supervised learning to semi-supervised and self-supervised learning to address the challenges of acquiring labels in the era of big data, the application challenges that arise in the era of big data have prompted continuous exploration in MVL, which involves leveraging advanced deep learning techniques e.g. representation learning,\textsuperscript{148} embedding learning,\textsuperscript{149} semi-supervised learning,\textsuperscript{150} and unsupervised learning\textsuperscript{151} to meet the feasibility requirements of practical applications.

For instance, Mao et al.\textsuperscript{152} employed a variational sparse Gaussian process model to handle multi-view data, modeling each view individually and achieving state-of-the-art recognition accuracy in kernel-based multi-view classification tasks. Subsequently, Sun et al.\textsuperscript{148} introduced representation learning into the multi-view data processing, further improving the classification performance of the Gaussian process model on multi-view datasets. Considering the challenge of acquiring labeled data, Tao et al.\textsuperscript{150} proposed a multi-view semi-supervised classification algorithm based on adaptive regression, which automatically balances the contributions of different views by optimizing adaptive weight coefficients, enabling the application of multi-view learning to large-scale datasets with incomplete labels. In addition to supervised and semi-supervised tasks that require labeled data, Ref.\textsuperscript{151} addressed the problem of multi-view unsupervised feature selection, and proposed a multi-view unsupervised feature selection algorithm with adaptive similarity and view weights, aiming to overcome the issue of neglecting the fundamental common structures across different views in traditional MVL methods.

(6) Transfer learning:

Although machine learning has achieved tremendous success, there are still limitations in certain scenarios. The underlying assumption of AI is that the training and test data follow the same distribution. Thus, the ideal scenario for machine learning is to have a large number of labeled training instances that share the same distribution as the test data. However, in many cases, collecting a sufficient amount of training data is often expensive, time-consuming, or even impractical. While semi-supervised learning can partially alleviate this problem by relaxing the requirement for a large amount of labeled data, in many situations, obtaining unlabeled instances is also challenging, leading to a significant decrease in accuracy or even failure of traditional machine learning or deep learning models.\textsuperscript{153}

Transfer learning aims to enhance the performance of models in the target domain by transferring knowledge from different but related source domains, reducing the dependence on a large amount of target domain data for constructing the target learner.\textsuperscript{154} Therefore, transfer learning with a focus on cross-domain knowledge transfer provides an effective approach to address the aforementioned problem. For example, Wan et al.\textsuperscript{155} proposed a deep transfer learning-based multi-source data fusion for UAV swarm radar signals algorithm to tackle the challenge of interference and pulse loss caused by pre-selection biases in the traditional pre-sorting step of radar signal ordering, particularly when the data in the target domain is not sufficiently abundant.

To facilitate the precise definition, we first provide mathematical definitions for the concepts of domain (Definition 5) and task (Definition 6) in the context of transfer learning, where the model outputs the conditional distribution of predictions for instances, we define \( f(X_i) \) in Definition 5 as \( f(X_i) = \{ P(\{x_k | x_i\} | y_k \in Y, k = 1, 2, \ldots, |Y|) \} \). In practical applications, a domain is often represented by multiple instances, some with labeled information and some without. For example, the source domain \( D_S \) corresponding to the source task \( T_S \) is typically represented as \( D_S = \{(x, y) | x_i \in X_S, y_i \in Y_S, i = 1, 2, \ldots, n_S\} \), where \( X_S \) represents the feature space of the source domain and \( Y_S \) represents the label set of the source task. The target domain consists of many unlabeled instances and/or a limited number of labeled instances.

**Definition 5.** (Domain) A domain \( D = \{X, P(X)\} \) in transfer learning consists of a feature space \( X \) and a marginal distribution \( P(X) \). Here, the symbol \( X \) represents a set of instances, i.e., \( X = \{x_i | x_i \in X, i = 1, 2, \ldots, n\} \).

**Definition 6.** (Task) A task \( T = \{Y, f\} \) in transfer learning consists of a label space \( Y \) and a decision function \( f \). The decision function \( f \) is an implicit function that learns the expected distribution of the data from the sample data.

According to the extended definition of transfer learning tasks in Ref.\textsuperscript{156}, Definition 7 covers the majority of scenarios in transfer learning. Here, \( m_T \) represents the number of transfer learning tasks, and if \( m_S = 1 \), the scenario is referred to as single-source transfer learning, whereas multiple-source transfer learning is referred to when \( m_S \) is greater than 1. Some studies focus on the setting where \( m_T \geq 2 \), e.g. Ref.\textsuperscript{157}. However, current research on transfer learning for MSIF tasks typically emphasizes scenarios where \( m_S = 1 \) (especially when \( m_S = m_T \)).

**Definition 7.** (Transfer learning) Given one or more source domains \( m_S \in \mathbb{N}^* \) with their corresponding source tasks \( \{(D_{S_i}, T_{S_i}) | i = 1, 2, \ldots, m_S\} \), as well as related target domains \( m_T \in \mathbb{N}^* \) with their target tasks \( \{(D_{T_i}, T_{T_i}) | i = 1, 2, \ldots, m_T\} \), transfer learning leverages the implicit knowledge from the source domains to enhance the performance of the decision function \( f(x_i) = 1, 2, \ldots, m_T \) on the target domains.

Roughly speaking, transfer learning can be categorized into two types based on the differences between domains: homogeneous transfer learning and heterogeneous transfer learning.\textsuperscript{158} To handle the case where domains have the same feature space, some scholars have proposed homogeneous transfer learning methods, and certain studies assume that the domains differ only in their marginal distributions, which allows for domain adjustment through methods such as sample selection bias correction\textsuperscript{159} or covariate shift adjustment.\textsuperscript{160} However, this assumption does not hold in many cases. For example, in sentiment classification tasks, a word may have different connotations in different domains, giving rise to word ambiguity.\textsuperscript{161} To address this issue, some research further adjusts the conditional distributions, i.e., the knowledge transfer process when domains have different feature spaces, resulting in heterogeneous transfer learning. In addition to distribution adaptation,
heterogeneous transfer learning also requires feature space adaptation, resulting in more complexity than homogeneous transfer learning.\textsuperscript{162}

(7) Brief analysis of AI related methods for MSIF

Based on the way that feature information or decision information extracted by the base models is processed during information fusion, AI-based MSIF can be classified into two architectures: joint fusion and collaborative fusion.\textsuperscript{163} The joint fusion architecture maps different source inputs to the same latent feature space through model mapping to facilitate the fusion of feature information. On the other hand, the collaborative fusion architecture achieves feature fusion by finding the correlation between the features or decision information of different source data in the latent feature space used for transition. The former enables information interaction among different source data through the same latent feature space, facilitating the complementarity of multi-source information. The latter, while preserving the unique and exclusive features of different source data, promotes mutual collaboration of information under certain constraints.\textsuperscript{164}

Based on the type of input data, fusion tasks can be divided into two categories: single-modal information fusion and multi-modal information fusion. The input data can take various forms such as text, image, speech, video, temporal signals, and discrete statistical features. For single-modal information fusion tasks, whether using traditional machine learning methods, deep learning methods, or other approaches such as reinforcement learning or MVL, it is possible to process a specific type of data using a single model. On the other hand, multi-modal information fusion often requires multiple models of different types to process different types of data, to ensure effective feature extraction.\textsuperscript{165} In specific tasks, traditional machine learning models are commonly used for low-dimensional data processing due to their simplicity and the need to avoid the “curse of dimensionality”. CNNs are proficient in handling data that emphasizes spatial relationships, such as images and videos. RNNs are suitable for modeling long-range dependencies in sequential data, e.g. text and audio. Transformers have the ability to model global relationships, making them effective in both two-dimensional data, such as images and videos, and one-dimensional data, such as text and audio, although they require a larger amount of data for training. In other words, traditional machine learning models are typically employed for data fusion tasks with low non-linearity, while complex tasks require the strong feature extraction capabilities of deep learning models, necessitating the selection of specific models based on the task at hand. In MSIF, the integration of multi-modal heterogeneous data is an unavoidable requirement, which calls for the fusion of multiple types of models. For example, considering that DenseNet cannot directly use text data as input, so traditional pre-training and transfer learning methods are difficult to deal with multimodal information fusion tasks involving image text recognition, Ref.\textsuperscript{166} proposed a multimodal information injection plug-in, which integrates text features into different visual channels based on each single word, and integrates channel visual features into different text words.

Despite the remarkable achievements of deep learning-based MSIF in various domains such as video classification,\textsuperscript{167} event detection,\textsuperscript{168} emotion recognition,\textsuperscript{169} speech recognition,\textsuperscript{170} and human behavior recognition,\textsuperscript{29} Section 3 provides a more detailed introduction to the corresponding application areas. However, compared to traditional machine learning models, deep learning models often require a large amount of training data to obtain effective weights for subsequent recognition or feature extraction tasks. In particular, Transformers typically require millions of training samples to achieve good generalization performance.\textsuperscript{171} and this limitation to some extent restricts the application domains of general deep learning models. To address this challenge, transfer learning adopts a strategy of training in the source domain and then transferring knowledge using a small amount of target domain samples when obtaining a sufficient number of training samples is difficult. This significantly reduces the sample requirements of the model,\textsuperscript{172} enabling deep learning models to be applied to small-scale datasets.

3. Applications

This section aims to provide an analysis of various application domains for MSIF to assist readers interested in specific application areas in finding relevant literature. In specific tasks, MSIF has shown remarkable achievements in numerous application domains by integrating global information, enhancing information accuracy and reliability, and eliminating redundant computations (Fig. 10).

Admittedly, due to limitations such as space constraints, it is not possible to enumerate all application scenarios and literature reports in this section; thereby, we select a subset of common scenarios for overview via the compilation of relevant literature, with the hope of helping researchers in related fields gain a quick understanding of the necessity and potential value of these technologies in practical applications.

3.1. Military applications

The battlefield information is dynamic, chaotic, and characterized by rapid changes, uncertainties, and information overload that soldiers face in combat environments. Consequently, modern military information systems face the challenge of providing comprehensive and accurate real-time information to the soldiers, enabling the military to take timely and effective actions.\textsuperscript{173} Due to the enormous amount of information available for command decision-making, commanders find it difficult to perform comprehensive analysis of raw information manually for battlefield situational assessment,\textsuperscript{174} such as analyzing and predicting enemy behavior and intent.\textsuperscript{175} Therefore, automated information processing methods based on MSIF play an indispensable and crucial role in modern warfare systems.

3.1.1. Threat detection

Detecting and preventing suicide bombings or attacks involving improvised explosive devices has always been a priority for military and other government organizations. One element in thwarting such
attacks is the observation of suspicious individuals through various means over time, including the use of scanners and other devices, travel records, behavioral observations, and intelligence sources. However, these types of data are highly complex and often prone to qualitative, subjective, and ambiguous situations. Sometimes, contradictory or deceptive information may also be present. 

Ergo, in counterterrorism, intelligence, and law enforcement scenarios, the fusion of uncertain multi-source sensitive information to form realistic, discriminative, and convergent assessments is a crucial safeguard for ensuring accurate decision-making by the system, which helps improve the efficiency of terrorist surveillance, reduce false alarms, and exonerate individuals who are improperly suspected.

Furthermore, MSIF can enhance the response speed of decision systems in detecting real threat targets, making it highly valuable in security monitoring tasks. For instance, Kessel utilized MSIF to improve decision reliability in a maritime surveillance and counterterrorism abnormal activity detection system, while Nowak et al. incorporated MSIF into embassy security decision systems.

3.1.2. Ground-to-Air surveillance
In theory, radar can detect any flying object with reflective signals. However, for certain stealth objects or objects flying beyond the coverage range of radar, it is difficult to achieve effective, comprehensive, and accurate air and ground observation solely relying on a single radar signal. As a result, some flying targets can fly at low altitudes and high speeds, and even possess the capability to deploy lethal weapons within a short period of time. Currently, there have been some advancements in Ground-to-Air surveillance methods for common aircraft detection. However, challenges such as low automation, communication delays, and vulnerability to deception in single-source signals still exist, leading to the continued reliance on manual visual observation in critical locations.

The fusion of multiple homogeneous radar signals in MSIF can effectively expand radar coverage, thereby obtaining more comprehensive air and ground observation results. On the other hand, the fusion of heterogeneous multi-source information, including radar, multispectral, and infrared, can effectively mitigate the vulnerability to deception in single-source signals and address the problem of “visibility” of detected objects in the surveillance airspace. Therefore, Ground-to-Air surveillance based on MSIF has become an important technological approach to the assurance of national airspace security.

3.1.3. Cognitive assistance
The battlefield environment is characterized by high complexity and time sensitivity, bombarding human operators with a vast amount of information, which can lead to information overload and increase the pressure on operators, potentially resulting in errors. Thus, automated information systems designed for military applications must have an effective user interface that focuses the information on the current task, reducing the information load for human operators. However, research in cognitive science and cognitive engineering has shown that humans are an essential part of any decision support design process. Excluding humans from the process can lead to mistrust between humans and machines, lack of preparedness when the machine fails, and ultimately, a decrease in overall system performance.

MSIF can effectively interpret a large amount of battlefield information and help human operators quickly understand the current situation in high-pressure environments through rational combination and presentation of information, which enables operators to make decisions with the assistance of efficiently presented information, while avoiding a crisis of human-machine trust caused by the system making decisions on their behalf. One typical application scenario of military cognitive assistance is pilot target identification based on multi-source image fusion. For example, in Ref., information from visible light and infrared images is fused to assist pilots in interpreting multimodal images under high pressure, enabling faster and more accurate target identification by fighter pilots.

3.1.4. Remote sensing and navigation
Target tracking based on MSIF is an important approach and objective task in various fields such as autonomous robotics, military applications, and mobile systems, which helps in collecting more accurate tracking information, such as improving tracking data in quality measurement scenarios to monitor the driver’s current status information, and tracking and predicting human action sequences, and other tasks.

In addition to civilian and commercial applications and research in these fields, target tracking and navigation techniques based on MSIF have been a focus of research in the military domain. For example, Yaakov Bar-Shalom has proposed several target tracking algorithms based on Bayesian estimation theory using MSIF, addressing a series of tracking and navigation requirements of the U.S. military. Numerous related papers have been published to report the advantages of MSIF technology in remote sensing and navigation domains.

3.1.5. Soldier status monitoring
A powerful military sensing system should process raw data in a multi-source fusion manner, encompassing classification, tracking, decision-making, prediction, and optimization. Soldiers, as crucial components of the Internet of Things, serve as the most flexible information nodes on the battlefield. The integration of heterogeneous sensors into soldier equipment provides multidimensional battlefield information to command centers. Moreover, as an interdependent and interconnected entity, soldiers can continuously communicate, coordinate, and collaboratively plan and execute tasks using wearable sensor networks.

Wearable sensor networks are fundamental elements of military smart devices, equipped with information-responsive sensing modules capable of collecting battlefield information under strict resource constraints. They facilitate data transmission and analysis in collaboration with other integrated modules within the devices. As a part of the modern Internet of Things, wearable sensor networks enable information exchange between humans and objects, enabling monitoring, analysis, and control of human bodies and surrounding environments.

In recent years, researchers have proposed specific solutions for optimizing sensor networks in distributed settings, decision fusion optimization, data redundancy handling, workflow scheduling. Overall, MSIF serves as the optimal approach to broaden the functional boundaries and enhance the network resilience of wearable sensor networks. Therefore, designing high-performance fusion methods is a crucial pathway for developing soldier state monitoring technologies based on wearable sensor networks.

3.1.6. Specific transmitter identification
The primary task of specific emitter identification is to recognize radar emitter sources. Radar plays a vital role in contemporary military affairs. However, due to the coupling of numerous different
emitter signals, both military and civilian electromagnetic environments are highly complex, and signals from the same emitter can interfere with each other, resulting in noisy and challenging-to-measure conditions.\textsuperscript{195} Consequently, distinguishing correct information from acquired signals and obtaining reliable results posed a challenging problem, making specific emitter identification a crucial safeguard for military information security and reliability.\textsuperscript{196}

MSIF has been an important technological approach to achieve specific emitter identification. Ref.\textsuperscript{197} proposed a quantum mechanics-based approach using the DST (see Section 2.3.1) to fuse multiple sources of information while considering the radar’s operational performance to simulate the reliability of identification results. Subsequently, Ref.\textsuperscript{198} further integrated spatiotemporal domain information based on the DST theory, combining the evidence measured by correlation coefficients with the influence of radar’s own performance in the process of MSIF, which fully considers the impact of time on fusion results, enhancing the dynamic nature of the outcomes.

3.2. Application of MSIF in nonmilitary scenarios

3.2.1. Human action recognition

Human action recognition has significant practical value and research significance in various fields such as ubiquitous computing, health monitoring, assisted living for the elderly, and sports activity monitoring, especially in areas such as behavior analysis, environment-assisted living, and self-monitoring in smart home environments.\textsuperscript{199} With the increasing aging population, there is a high risk of falls among the middle-aged and elderly population. Recognising the occurrence of actual falls can help prevent negative health cost trends in this population, which highlights the important role of fall detection and pose recognition in human action recognition. Furthermore, in certain scenarios, human action recognition requires the detection and recognition of physiological signals, which enables real-time interaction with related medical diagnoses, allowing individuals to monitor their own health status and achieve the goals of early detection, prevention, and treatment.\textsuperscript{200}

Early research in human action recognition primarily focused on using single-modal sensors,\textsuperscript{201} single-sensor features,\textsuperscript{202} and classifiers.\textsuperscript{203} While these detection methods are relatively simple, they sometimes struggle to effectively differentiate complex activity details, and single-sensor data is susceptible to data uncertainty and indirect data acquisition influences.\textsuperscript{204} Therefore, in order to fully utilize data, features, and classifiers for effective health and activity monitoring, fusion strategies are needed to enhance the reliability, robustness, and generalization capabilities of recognition systems.\textsuperscript{205} Some researchers adopt the approach of combining multiple homogeneous or heterogeneous weak classifiers (decision-level fusion, see Fig. 7(c)) to improve the robustness, accuracy, and generalization of classifiers.\textsuperscript{206} By fusing the outputs generated by different classification models, uncertainties and ambiguities can be reduced, achieving higher performance that is difficult to attain when using classifiers individually.\textsuperscript{207}

3.2.2. Stock price forecast

Traditional methods for stock market prediction typically use historical stock datasets to forecast stock price fluctuations.\textsuperscript{208} But these approaches overlook the rich information available from other sources e.g. the internet, databases, chat logs, emails, and social networking sites\textsuperscript{131}, resulting in unsatisfactory prediction outcomes. Information related to stocks can be classified into two categories: quantitative numerical data and qualitative textual data. Quantitative data includes historical stock prices and economic indica-
By integrating all the information received from onboard sensors with information from neighboring vehicles via multi-source fusion, connected vehicles can achieve more accurate and comprehensive situational awareness, thereby providing more precise positioning for intelligent driving to assist drivers and avoid catastrophic consequences.  

4. Challenges and future

This section provides a brief overview of the current challenges in MSIF and identifies key issues that need to be addressed in theory and applications. Then, the future directions of MSIF are discussed, based on these issues.

4.1. Key issues and challenges

4.1.1. Data level

The challenges in MSIF arise primarily from the problem of complete data, where the data provided by sensors is prone to certain levels of inaccuracy and measurement uncertainty. Therefore, effective algorithms for MSIF should be able to express these imperfections and utilize the redundant information across multiple data sources to mitigate the impact of incomplete data on final decision-making. Specifically, there are five categories of issues caused by incomplete data in MSIF systems:

1. Uncertainty of information. Uncertainty of information refers to the lack of judgment and absolute certainty in the observer’s perception of objective phenomena.  

The uncertainty of information stems not only from imprecision and noise in measurements but also from the ambiguity and inconsistency present in the environment, making it difficult or even impossible to accurately differentiate observation objects based on the available information. For example, in image classification tasks, similar color and texture information may result in lower recognition accuracy for most traditional methods due to the greater uncertainty of information compared to deep learning algorithms, which reflects the superiority of AI in certain specific tasks and underscores the significant impact of information uncertainty on decision outcomes.

2. Conflicting data (source conflict). To fully utilize the data, it is necessary to acquire unified-format data from multiple sources. However, due to issues such as duplication, semantic diversity, and data quality among heterogeneous knowledge sources, conflicting data may arise, which is particularly relevant when fusion systems rely on combination rules based on belief functions. Therefore, when performing MSIF, it is necessary to conduct operations such as conflict detection, entity disambiguation, and entity alignment on the data.

3. Missing information. In the era of the Internet of Things, the monitoring of massive data has become a necessary and routine task. Dealing with massive monitoring data not only consumes significant computational resources but can also impede the efficiency of information transmission in the entire information network, leading to information loss and reduced or even ineffective decision accuracy of fusion algorithms.

4. Data heterogeneity. Data heterogeneity is evident in structured, semi-structured, and unstructured data. Traditional MSIF mainly focuses on the fusion of structured data. However, with the advancement of sensor technologies, there is a growing need to fuse more unstructured data, each with its own feature representation. For instance, text data is typically represented by discrete word vector features, while images are represented in the form of three-dimensional tensors. Therefore, data heterogeneity poses a significant challenge that must be addressed in MSIF.

5. Data correlation: Data correlation refers to the interference from external information sources that impacts the measurement process of sensors, resulting in similar biases in data distributions, which is particularly common in distributed fusion systems. For example, some sensor nodes in wireless sensor networks may be exposed to the same external noise, leading to similar biases in their measured values. If the correlation among such data is not considered to remove the common external noise interference, fusion algorithms may yield low confidence in decision outcomes due to this “unified interference”.

In addition to the five key issues caused by incomplete data mentioned above, unlike individual signal processing tasks, MSIF often faces challenges, such as data alignment, fusion processing of dynamic data (or static data accompanied by some dynamic components), and source state evaluation.

The process of data alignment, also commonly referred to as sensor data registration, involves transforming sensor data from each sensor’s edge terminal data to a common frame of reference before information fusion processing, and addresses calibration errors caused by individual sensor nodes. In both homogeneous and heterogeneous MSIF systems, the operating frequencies of sensors may vary, requiring effective fusion methods to handle information at multiple time scales to accommodate such temporal variations in the data. This becomes especially crucial in distributed fusion environments, where different data components may follow different paths before reaching the fusion center, resulting in out-of-order data arrival. Achieving data alignment in real-time applications effectively mitigates the degradation of system performance caused by temporal disorder in information processing. Therefore, at the data level, data alignment becomes a critical issue that cannot be overlooked in MSIF systems.

The data collected in practical applications is typically not limited to static forms but comprises multiple sources of dynamic data, particularly interval data used to describe dynamic phenomena such as temperature variations, stock price fluctuations, and blood pressure. MSIF systems obtain interval data from various locations or sources, leading to observations that can be either time-invariant or time-varying. Furthermore, due to the significant impact of information freshness (i.e., the speed at which data sources capture changes and update accordingly) on the effectiveness of decision outcomes, it is necessary for MSIF algorithms to dynamically fuse recent measurement histories in the data processing. Therefore, effectively handling and fusing these dynamic or dynamically-static combined data pose important challenges with practical application value.

4.1.2. Application level

In the early 21st century, MSIF faced various challenges in practical applications, including limitations in theoretical techniques and hardware devices, sensor array design, algorithm selection, system evaluation, and poor human-computer interaction. Fortunately, with the rapid development of technological technologies, e.g., AI, as well as significant improvements in hardware performance in recent years in the fields of computer science and engineering, these issues have been effectively addressed. Nevertheless, along with the advent of the era of massive information and the Internet of Things brought about by new technologies, a series of new challenges have emerged for MSIF. We categorizes these challenges into three types: general MSIF, real-time MSIF, and event-driven MSIF. Despite the improvements in technology and hardware, these challenges pose new obstacles that need to be addressed in the field of MSIF.

1. General MSIF

For general MSIF applications, two challenges are faced in
practical scenarios: data association and data dimensionality.

Data association involves reordering a set of observations of detection objects to describe the true state of the targets, and accurate data association is crucial for ensuring the correctness of subsequent information processing. Data association is typically performed before state estimation and is commonly used in multi-target tracking tasks. Compared to single-target tracking cases, multi-target tracking introduces significant complexity to fusion systems. In multi-target tracking tasks, data association problems exist in two forms: measurement-to-track association and track-to-track association. The former refers to the problem of identifying which target (if any) each measurement originates from, while the latter deals with distinguishing and combining tracks, which estimate the state of the same real-world target.

The challenge brought by data dimensionality mainly stems from the data dimension reduction process in the pre-processing stage before information fusion. That is, under the condition of allowing a certain degree of compression loss in the collected information, the measurement data is pre-processed at the edge terminals of each sensor node or at the fusion center to compress it into lower-dimensional data. In edge terminal pre-processing, this stage can save the communication bandwidth and power required for transmitting data. On the other hand, when global pre-processing is performed at the fusion center, it can reduce the computational load of the central fusion node. Therefore, for practical MSIF, whether deployed in edge intelligence or central processing, optimizing data dimensionality is a valuable challenge.

(2) Real-time MSIF

With the rapid development of methods and technologies such as AI that possess efficient and fast data processing capabilities, the systems of MSIF have been applied in numerous short-cycle data analysis and real-time information fusion tasks. Real-time MSIF aims to quickly detect events to complete the description or analysis tasks of targets continuously acquire new information from a large amount of data generated by multiple signal sources, and efficiently and timely process multi-source information. Therefore, real-time MSIF support systems continuously obtain information from new data for information fusion, improve the accuracy of information fusion, and better support real-time processing and execution of local operations, thereby facilitating cloud data collection and supporting big data analysis for cloud applications.

Although many real-time MSIF methods have been proposed to effectively handle incremental data in the real world, real-time MSIF technologies still need to address key issues such as dynamic data processing requirements, reduced time costs, reliability of data transmission (selecting effective methods to reduce delays based on computational complexity and execution time), and uncertain data fusion.

(A) Requirements of dynamic data processing. In complex and changing scenarios, the environment evolves over time (for example, in autonomous driving tasks, the movement of users and vehicles forms a dynamic data stream, causing the collected data to change over time), thereby imposing dynamic information processing requirements on algorithms.

(B) Reduced time costs. In the real-time fusion process, a large number of sensors continuously generate and transmit data, leading to the generation of a significant amount of redundant and erroneous information, resulting in high computational and application costs for data fusion.

(C) Reliability of data transmission. The non-deterministic delays and loss of flow during data transmission make it difficult to achieve large-scale real-time MSIF within the specified time, leading to a degradation in system performance.

(3) Event-driven MSIF

Event-driven MSIF is primarily used in resource-constrained applications such as efficient target detection tasks and industrial process monitoring. In this context, “event” refers to changes or transitions in states, where the fusion system starts collecting various sensor data and performing aggregation and fusion only when the entity or entity attribute undergoes a change that satisfies the event triggering rules, in order to accomplish predefined tasks. Event-driven MSIF technology has been widely applied in various fields due to its advantages of reducing unnecessary data transmission on the network, saving computational resources and network bandwidth, improving resource utilization efficiency, and prolonging system lifespan.

However, in actual engineering, event-driven MSIF still needs to address three key challenges: the effectiveness of event-driven triggering strategies (event-driven information fusion technology enhances its usability by establishing different event triggering strategies), the rapidity of event response (the ability to timely respond to corresponding events is crucial for event-driven fusion systems to handle dynamically changing tasks), and the temporal characteristics of events (almost all real-time application systems exhibit temporal characteristics, which broaden the application prospects of studying transient behaviors within a limited time range compared to steady-state performance under normal conditions).

4.2. Hot topics in current research

4.2.1. Technical difficulties in current theoretical research

Although the theory and techniques of MSIF have made significant progress and development, there is a continuous need to expand and advance MSIF at both the theoretical and applied levels to meet the growing real-world demands resulting from the emergence of new application domains. In terms of theoretical research, the integration of different fusion methods, standardization of fusion approaches, and evaluation of fusion results have always been the focal points and technical challenges in the field of MSIF.

(A) Integration of different fusion methods. Current fusion methods aim to integrate low-level attributes into higher-level attributes to increase generality and improve the portability of results. However, these methods are relatively independent, and few reports have used more than one fusion method for multi-source data processing in a single study. For example, Zhang et al. attempted to perform multi-method fusion by combining the DST theory and clustering techniques, but they did not achieve the integration of multiple fusion methods.

(B) Standardization of fusion methods. To date, researchers have classified relevant papers on MSIF based on data types, architecture types, and fusion levels, but there is no standardized classification method for fusion methods. As a future research direction, analyzing, characterizing, formalizing, and clustering literature from different perspectives would be interesting and meaningful. Standardized research would also promote the development of the field and provide clear guidance for future researchers in defining new techniques of MSIF.

(C) Evaluation of Fusion Results. The idea of improving information quality using MSIF is significant in various fields such as data analysis (requiring reliable descriptions or supporting
information, AI (demanding input information without quality loss), and network security (detecting message patterns and identifying network attacks requiring reliable input information). Therefore, it is necessary to quantify the quality of fused information for measuring the credibility of information after MSIF in different scenarios.

In addition to the above theoretical research directions, this section explores and discusses the future research directions of MSIF that have significant scientific and practical value, based on current trends and the upgrading application demands brought by related technological developments, and examines the technical approaches, application directions, and development prospects to outline the promising avenues for future research in MSIF.

4.2.2. Technology based on AI
In recent years, more than two-thirds of MSIF work has been accomplished through the implementation of AI, and this proportion is steadily increasing (see Fig. 11). The reason for this is that AI has greatly expanded the application scenarios of MSIF. Furthermore, the growing demands for practical applications in new tasks have further driven AI development, creating a mutually beneficial relationship between the two. Therefore, research on the AI-based theory has become an important direction for the future development of MSIF at the theoretical level.

![Number of publications](image)

Fig. 11 Promotion effect of AI on MSIF in academic field, in the past five years. (Note: data is sourced from Web of Science, and only non-review reports are included)

Despite the significant achievements and substantial theoretical accumulation in the field of AI, thanks to the diligent efforts of numerous scholars and practitioners, there are still some theoretical and technical aspects that require further research and development for algorithms specifically designed for MSIF tasks. These areas include but are not limited to federated learning, attention mechanisms, lifelong learning, etc.

(1) **Federated learning** is a novel learning paradigm that decouples data collection and model training through multi-party computation and model aggregation. As a flexible learning environment, federated learning has the potential to integrate with other learning frameworks. Recent research reports have presented a series of successful cases of federated X learning, where X can be multi-task learning, meta-learning, transfer learning, unsupervised learning, and reinforcement learning, among others.

(2) **Attention mechanism** is one of the most important concepts in the current field of deep learning, which dynamically adjusts the focus of a neural network by scaling the features with corresponding weights calculated by independent attention modules. The inspiration for the attention mechanism comes from the human biological system’s ability to focus on specific features when processing a large amount of information, and this ability to selectively process specific information in massive data is expected to serve as a theoretical foundation for addressing the increasing challenge of handling the growing information volume in MSIF.

(3) **Lifelong learning** refers to the ability of AI to acquire new skills without compromising old ones to adapt to task changes, which involves applying previously learned knowledge to new tasks while preserving limited resources such as computing power, memory, and energy. Lifelong learning is essential in scenarios where resources are constrained, and adaptability to changing tasks is required.

Furthermore, in recent discourses, the inherent causal logical properties carried by probability theory have effectively propelled the advancement of various subdomains within deep learning, including studies on interpretability, security, etc. However, the integration of traditional probability theory with deep learning still faces several limitations, as evidenced by current reports. For instance, in terms of result diversity and reliability, if all predictive outcomes of a deep learning model satisfy the statistical expectations of a probability distribution, the model becomes susceptible to deception from adversarial attacks crafted based on that sampling distribution. Consequently, within the era where AI, primarily driven by deep learning, is rapidly propelling the evolution of MSIF, how to balance the advantages of probability theory and ensure the robustness, reliability, security, and accuracy of MSIF models based on deep learning emerges as a highly significant avenue of research, which is to develop deep-learning-based MSIF under generalized probability theories.

4.2.3. Virtual-real fusion
With the increasing demand for industrial Internet, digital twinning, and big data technologies, there are a tremendous number of potential application scenarios throughout the product lifecycle. Driven by this trend, discrete manufacturing enterprises urgently need intelligent upgrades. To facilitate the development of the discrete manufacturing industry, many countries have formulated new national advanced manufacturing strategies. Examples include Germany’s Industry 4.0, which revolves around Cyber-Physical Systems, and China Manufacturing 2025.

Currently, the large-scale customization development of discrete manufacturing enterprises requires a high level of manufacturing flexibility to ensure quality assurance and timely delivery for all production orders. Additionally, to mitigate uncertainties in the manufacturing process, such as machine failures, processing time, and product demands, it is necessary to evaluate multiple design variables, including equipment selection, control schemes, workpiece development, and process and service design solutions. In the discrete manufacturing industry, a better balance among quality, efficiency, and cost is needed, and performance indicators related to materials, energy, and waste must be designed and analyzed more rationally.

How can the cost of producing prototypes and testing them be reduced? How can extreme testing be conducted on prototypes that cannot be tested in a laboratory? How can all the information and results from these tests be assimilated to provide accurate predictions of future behavior? How can physical assets be monitored in real-time, with alerts received before any serious errors occur? And how can humans access real-time information about all the components involved in physical assets and make meaningful real-time analysis of this information to make timely, robust, and efficient decisions based on their future operations? The fusion of the real
physical and virtual worlds, virtual-real fusion, provides a highly feasible answer to these questions. Specifically, in the physical realm, communication and connectivity between hardware devices are achieved through field buses, Ethernet, gateways, and communication protocols. Simultaneously, through edge access technologies such as the Internet of Things and information system interfaces, multiple sources of information in the physical space (people, machines, materials, methods, environment, measurements, and other production elements and processes) are integrated into the virtual space. By using MSIF to create digital twins, deeper integration, interaction, and control of the physical and virtual worlds can be achieved. Solutions in the physical space can be simulated and optimized through multiple iterations in the virtual space, which is where the significant value of the fusion of the physical and virtual worlds lies. In other words, optimized, improved, and validated instructions from the virtual space are applied to control the physical space, enhancing the feasibility of instruction execution. The execution results and information from the physical space are continuously transmitted to the virtual space for information synchronization and continuous analysis.

Driven by a series of real-world demands such as industrial intelligence upgrading, the virtual-real fusion has achieved remarkable results in many application scenarios. For example, to address the challenge of data-driven model performance limited by training data, Ref. developed an energy estimation model for online building control and optimization using integrated system identification and MSIF methods for the evaluation and development of virtual and physical buildings. Ref. constructed a multi-modal sensing and navigation fusion laboratory based on MSIF. By integrating multi-modal perception algorithms, this laboratory can understand users’ real intentions, enhance human-machine interaction efficiency, and avoid resource waste and potential risks in physical experiments through virtual experimentation, thus improving users’ operational experience and sense of reality. In the field of industrial equipment maintenance, Shen et al. employed the fusion of depth and color images to build a virtual-real fusion maintenance testing scenario. Ref. combined MSIF and digital twin technologies to replicate the real tunnel scene in the virtual space, providing new means for the digital maintenance of tunnels, addressing issues such as fragmented videos, separation of videos and business data, and the lack of 2D and 3D interactive response in tunnel digital operations.

Despite these achievements, challenges remain in achieving a virtual-real fusion on the “brain-in-a-vat” level, limited by the current level of development in information acquisition, fusion schemes, model construction, and hardware performance. These challenges include low efficiency in human-machine interaction, weak sense of reality, distortion of virtual world laws, and difficulties in modeling virtual time. To overcome these issues, MSIF has become a key technical approach.

4.2.4. Human-machine integration

With the advancement of technology, machines are becoming increasingly intelligent and gradually moving towards semi-autonomous and autonomous systems, thereby the harmonious coexistence of humans and machines is taking center stage in human history. MSIF is essentially the process of humans utilizing sensor devices deployed on relevant machines to aggregate information for subsequent tasks. Therefore, human-machine integration represents an advanced form of MSIF, and MSIF serves as a crucial guarantee for achieving human-machine integration.

Currently, human-machine integration can be categorized into different levels of fusion from three perspectives.

(1) From the perspectives of sensing, perception, and action

Information fusion: Machines obtain consistent or shared data with human senses (visual, auditory, tactile, olfactory, etc.) through sensors, e.g., optics, acceleration, pressure, and gas sensitivity sensors, which often requires careful analysis and processing to obtain, i.e., whether the information mined by the machine is the information that humans need.

Cognitive fusion: It examines whether human perspectives on things align with machine perspectives.

Thought fusion: This level represents a higher manifestation of human-machine coexistence. It explores whether there is consistency in thinking and ideology between humans and machines.

Behavior fusion: The purpose or goal of human-machine coexistence should be aligned. This level assesses whether the behaviors exhibited by humans and machines are consistent in order to achieve the common objective.

(2) From a security perspective

Essential security: This level focuses on designing production equipment or systems to have inherent safety features. Even in cases of operator error or equipment failure, they should not cause accidents.

Comfortable security: It pertains to ensuring that machines do not produce inappropriate sounds, say untimely words, or engage in actions that violate norms. These factors can evoke human discomfort and decrease the level of comfort and safety.

Privacy security: This level aims to protect three types of behaviors: personal information that individuals do not wish others to know or find inconvenient to know, personal affairs that individuals do not wish others to interfere with or find inconvenient to interfere with, and personal domains that individuals do not wish others to invade or find inconvenient to invade.

Information security: This level encompasses the confidentiality, authenticity, integrity, unauthorized copying of information, and the security of the systems where the information resides.

(3) From the perspective of coexistence forms

Human-led: Humans act as the primary agents of human-machine coexistence, responsible for guiding the coexistence goals and directions. They provide timely corrections to machines, review their actions, and make final decisions.

Machine-led: These machines serve as the primary agents of human-machine coexistence, while humans become integral parts of unmanned systems. Machines have the ability to autonomously make decisions by synthesizing various information.

Mixed-led: Humans and machines learn from each other, mutually enhancing their cognitive abilities and levels. They progress together through collaborative efforts.

Currently, the challenge of achieving harmonious coexistence and collaborative development between humans and intelligent machines or unmanned systems is of great significance. In the process of promoting the development of human-machine integration through MSIF, several research areas have shown immense potential that cannot be overlooked, which deserve particular attention as future research directions.

(1) Knowledge graph fusion

Knowledge graph representation is essentially a large-scale semantic network that aims to describe various entities and concepts existing in the real world, along with their relationships, in a form closer to human cognitive understanding, which has been widely applied in intelligent search, personalized recommendations, intelligent question answering, and other domains. In the era of big data, the presence of multiple stages makes the representation of knowledge graphs increasingly complex and massive, thus necessitating a reconsideration of the accuracy and efficiency of algorithms for
constructing multi-source knowledge fusion. Although parallel processing techniques have been employed in knowledge graph fusion inference to enhance model efficiency, the existing techniques predominantly focus on knowledge reasoning, which leaves many challenges to be addressed regarding the establishment of large-scale knowledge fusion frameworks. Language knowledge bases are becoming crucial sources of knowledge for both humans and AI applications. While the fusion techniques based on single graphs have achieved promising results, there is still much room to explore in cross-graph knowledge fusion applications.

Therefore, the fusion of multiple sources of information and graph-based representation in knowledge fusion presents a research direction with significant potential, which is expected to drive research and development in knowledge graphs in areas such as web-scale search and natural language processing, facilitate the construction of domain-specific knowledge graphs, and bring about substantial social and economic benefits. Moreover, the knowledge graph fusion serves as an important technical approach and guarantee for achieving “cognitive integration” between humans and machines.

(2) Multimodal fusion

Since 2010, AI has revolutionized fields such as speech recognition, image recognition, and natural language processing, which typically involve single-modal signals as input, while many applications in the field of AI unavoidably require joint processing of multiple modalities, especially in tasks involving MSIF. Furthermore, both humans and machines need to achieve accurate and comprehensive descriptions of the environment, situations, and decisions, which rely on the integrated analysis of different modalities of data.

Therefore, researching applications and theoretical techniques for multimodal signal fusion becomes essential in promoting the rapid development of AI in MSIF and achieving “cognitive integration” in the process of human-machine integration. On the other hand, emotion is a unique cognitive ability of humans, and communication between individuals involves rich emotions. High-level interpersonal fusion cannot be achieved without emotional fusion. In order to create a highly personified human-machine interaction environment, machines must understand multimodal data and generate multimodal emotional content that resonates with humans.

Fundamental research in this field not only helps us understand the mechanisms of cognitive intelligence but also has significant value for many real-world applications. Currently, the fusion of multiple emotions by machines faces three challenges: (A) perceiving and adjusting subtle expressions of emotions in different ways; (B) ensuring consistency and rationality of data across all modalities; (C) extracting core representations and approaches to capture latent and invariant emotions in multi-source information.

(3) Human-machine fusion for massive data.

In the era of big data, although AI based MSIF has successfully empowered numerous machines to enhance their intelligence, tasks involving massive data, such as stock price prediction, securities trading, and e-commerce, still face the formidable challenge of processing ultra-large-scale data in the field of human-machine interaction. Taking e-commerce as an example, this domain confronts challenges related to ultra-large-scale data and complex human-machine interactions within the entire retail chain, which requires large-scale, complex, task-oriented multimodal intelligent human-machine interaction technologies to serve hundreds of millions of users in a personalized and efficient manner.

Therefore, to facilitate the open-source and open-license frameworks for complex human-machine interaction systems, build large-scale datasets and algorithm validation platforms, and promote the development of intelligent MSIF, it is necessary to conduct in-depth research on intelligent human-machine integration under the context of ultra-large-scale and complex objectives.

(4) Multi-granularity information fusion

In information systems, information granularity is used to measure the degree of knowledge clustering in a domain, which indicates that coarser domain classifications result in larger information granularity, and vice versa. Traditional data mining and pattern recognition tasks always aim to analyze data at the finest granularity possible, while granular computing believes that solving different tasks using information at different granularities can achieve more accurate descriptions and higher spatiotemporal efficiency, which aligns well with human cognition and behavior.

In the field of cognition, humans use light-years to measure distances between stars and nanometers to describe distances between atoms, which is a way of selecting different granularity levels in space to obtain higher spatiotemporal efficiency in descriptions. Similarly, in the temporal domain, the narrative style of “November 15, 2022, 17:26:18” combines time information at different granularities to achieve a more precise description. The concept of multi-granularity control is abundant in human behavior as well. For example, when artists paint, they start with outlining and then adjust details, demonstrating different granular control of arm muscles. In both the perception of information and the control of behavior, humans and machines exhibit forms of multi-granularity information fusion. Therefore, whether aiming to achieve deeper levels of human-machine fusion or pursue higher-level MSIF, multi-granularity information fusion is a crucial technology with significant research value and practical significance in the future.

4.3. Perspective

Based on the development process, current research trends, and the trajectory of technological and theoretical advancements, this section provides a perspective on the future development of MSIF. In the future, MSIF are expected to make significant breakthroughs in areas such as consciousness fusion and autonomous intelligent fusion.

To achieve these breakthroughs, interdisciplinary collaboration is crucial: collaboration among researchers from fields such as cognitive science, AI, machine learning, and human-computer interaction can facilitate the integration of diverse expertise and perspectives. By pushing the boundaries of research and fostering interdisciplinary collaborations, researchers can unlock new frontiers in the field, leading to transformative applications in various domains.

4.3.1. Consciousness fusion

To advance the field of consciousness fusion, researchers can explore novel approaches to integrate information from diverse sources, including sensory inputs, cognitive processes, and contextual understanding. This integration can lead to a more comprehensive and holistic representation of human consciousness, bridging the gap between human and machine cognition.

Note: Given that consciousness fusion among humans necessitates careful consideration of underlying ethical factors, including privacy, intention, individuality, and personal independence, any potential approach to achieve consciousness fusion between humans must first be scrutinized from ethical, philosophical, and societal perspectives to assess potential risks. Therefore, this paper focuses solely on scenarios involving consciousness fusion between machines and does not address consciousness fusion among humans.
Consciousness fusion between machines: In the context of various task scenarios and experiences, both homogenous and heterogeneous machines with evolving autonomous learning capabilities may exhibit varying levels of intelligence. When these machines collaborate on a shared task, their differing levels of intelligence necessitate the integration of their cognitive processes, objectives, and decision outcomes to form a unified collective consciousness, enabling them to achieve higher levels of efficiency and intelligence.

Consciousness fusion between humans and machines refers to the integration of human consciousness with machine intelligence in a way that blurs the boundaries between humans and machines, enabling humans to effectively utilize machine intelligence for various tasks. Human-machine consciousness fusion takes on various forms, including but not limited to advanced applications of exoskeleton technology, which assist individuals with disabilities in enhancing their motor abilities, as well as utilizing machines as carriers of human consciousness to perform tasks in high-risk environments or engage in learning, training, and communication within virtual environments. In the future, human-machine consciousness fusion has the potential to fundamentally transform the human-machine relationship, making machines more intelligent, efficient, and human-centric, leading to greater well-being and progress in human society.

4.3.2. Autonomous intelligence fusion

With the advent of the information explosion era, the integration and analysis of heterogeneous, multisource information across domains, disciplines, and granularities have become increasingly prevalent. The long-term goal of MSIF is to achieve autonomous intelligence fusion of these diverse information sources. This section identifies four levels of autonomy: intelligentization of data, of models, of fusion methodologies, and of evaluation.

(1) The intelligentization of data refers to the notion that “data” is not merely simple and singular, which possesses characteristics and functionalities such as traceability, self-learning, self-purification, self-repair, proactiveness, and self-protection. In other words, intelligent data cells have built-in intelligent pre-processing capabilities.

(2) The intelligentization of models entails the ability of MSIF models to intelligently and autonomously adjust, update, and switch to adapt to dynamic changes in data sources, tasks, and conditions. This enhances the adaptability, robustness, and accuracy of fusion models.

(3) The intelligentization of fusion methodologies involves recommending fusion algorithms or approaches based on task objectives. Leveraging intelligent data and intelligent models, the fusion process incorporates functions such as information filtering, induction, classification, analysis, and reasoning, enabling autonomous decision-making.

(4) The intelligentization of fusion evaluation refers to the adaptive variation of evaluation criteria and requirements in response to dynamic changes in the external environment, conditions, and task objectives.

5. Conclusions

As a comprehensive discipline with interdisciplinary innovation potential and broad application prospects, MSIF has demonstrated outstanding performance and tremendous potential in fields such as robotics, intelligent system design, and pattern recognition, which has been a powerful driving force for academia and industry to develop more advanced fusion theories, methods, frameworks, and application approaches. To facilitate a quick understanding of the progress, latest achievements, and development trends in MSIF for interested readers, and to promote fruitful outcomes in this discipline, this paper presents a statistical analysis of academic reports primarily from international conferences and journals related to information fusion, and briefly describes the research overview of domestic and international scholars in terms of definition, theoretical methodologies, and applications. Additionally, we analyze the key challenges faced by MSIF and provide an outlook on future development directions.

Overall, the continuously emerging remarkable achievements are consistently proving the effectiveness and superiority of traditional MSIF in numerous application domains. Furthermore, AI has become a significant driving force for the rapid development of MSIF and will serve as the primary technical approach for the next generation of MSIF for a considerable period. While AI presents unprecedented development opportunities for MSIF, it also poses a series of urgent and daunting issues that need to be addressed, and these challenges, in turn, drive the continuous advancement of MSIF.

Acknowledgements

This study was co-supported by the National Natural Science Foundation of China (Nos. 62233003 and 62073072), the Key Projects of Key R&D Program of Jiangsu Province, China (Nos. BE2020006 and BE2020006-1), and the Shenzhen Science and Technology Program, China (Nos. JCYJ2021033122020005 and JCYJ20220818101206014).

References


73. Xiao F, Pedrycz W. Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 2022;45(2):2054-70.


Multi-source information fusion: Progress and future


Declaration of interests

☐ 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.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

None