

# Staying Stronger Together: Synergy From Multiple Sensors

Hsiao-Ying Lin, Huawei France

*Deep-learning-based multisensor fusion creates opportunities and challenges.*

**A** Japanese anime television series called *Neon Genesis Evangelion* features the Magi system, a supercomputer that makes decisions from three perspectives—that of a scientist, that of a mother, and that of a woman. The people in this series believe that a better decision can be made by considering a greater number of perspectives. Similarly, in the real world, people expect to obtain advanced functionalities with high accuracy and robustness by fusing greater amounts of sensory information. A biological example of multisensor fusion is the human brain, which fuses sensory information and makes inferences. Deep-learning-based multisensor fusion aims to approximate the mysterious working of the human brain and opens up an avenue for various applications, such as autonomous driving and robotics. In this context, this article



focuses on deep-learning-based multisensor fusion and discusses the associated opportunities and challenges.

According to a report published by the National Highway Traffic Safety Administration of the United States,<sup>1</sup> 6,516 pedestrians lost their lives in automobile-pedestrian crashes in that country in 2020. Advanced driver-assistance systems (ADASs), such as automotive collision-avoidance subsystems, can help to prevent such crashes and, possibly, save many lives. In such systems, multisensor fusion perception modules are important enablers. For example, a forward-collision warning system uses cameras and radar sensors to detect the potential for a crash and immediately warns the driver. Moreover, a few advanced forward-collision systems autonomously brake or steer the vehicle upon detecting the potential for a crash.

Robots working in hostile environments greatly reduce the number of human fatalities and injuries. Multisensor fusion modules are crucial to ensuring that robots can precisely perceive the physical world and moving objects. For example, firefighting robots can detect fires and implement countermeasures automatically or on the basis of remote instructions from human commanders while firefighters are kept at a safe distance. These firefighting



robots rely on infrared or ultraviolet detectors and visual cameras to understand their working environment.

## SENSING TECHNOLOGY IS A CORNERSTONE

Sensing technology pertains to the use of sensors for sensing physical environments and to the transformation of the sensed information into a readable format. There are two main categories of sensing technologies. The technologies under the first category mimic human senses, namely, touch, sight, hearing, smell, and taste. Image and acoustic sensing technologies have been studied extensively, but there is considerable room for exploring sensing technologies pertaining to touch, smell, and taste. The technologies under the second category address what lies beyond human senses, such as ultrasound, infrared signals, light detection and ranging (lidar), radar, and satellite signals.

In multisensor fusion, homogeneous or heterogeneous sensory information is combined to achieve the desired functionality. Multiple homogeneous sources of sensory information may yield extra information or offer a certain level of robustness through redundancy. For example, a stereo camera consisting of two camera devices not only captures images from different angles but also captures depth

information. A region of overlap exists in the pair of images captured using a stereo camera, and each camera can, thus, serve as the other's backup. In addition, multiple homogeneous sources of sensory information may be used in a complementary manner. For example, several cameras may be mounted on a driverless car to provide a 360° view of the car's surroundings.

Multiple heterogeneous sources of sensory information create diverse fusion opportunities with potentially high fusion complexity. For example, an automotive perception system may consist of a visual camera and a lidar sensor for detecting and locating objects on the road, although the images captured by the camera and the 3D cloud points provided by the lidar sensor must be synchronized before they are fused. Deep-learning-based heterogeneous sensor fusion is based on multimodal machine learning, in which data obtained from various sensor modalities are used to arrive at a final prediction. Heterogeneous sensor fusion has two main characteristics in the context of multimodal machine learning. First, it relies heavily on time synchronization among all sensed data, whereas multimodal machine learning may rely on semantic alignment. Second, heterogeneous sensor fusion uses signals from the physical environment, whereas

multimodal machine learning may use data from cyberspace.

Researchers have been working on specific hardware designs and signal processing techniques to advance sensing technologies that mimic human senses, such as touch and odor sensors, as well as superhuman senses, such as the pressure and temperature sensors used in robotic exoskeletons. The diversity of sensing technology serves as a strong support base for the evolution of multisensor fusion technology.

## DEEP-LEARNING-BASED MULTISENSOR FUSION STRATEGIES

Owing to the diversity of sensing technologies, various types of signals are represented using different data types, such as images, audio, video, radar point clouds, and lidar 3D point clouds. Three main types of data fusion strategies exist, which differ in terms of how data are fused, as illustrated in Figure 1.

First, in data-level fusion strategies, sensed data are integrated at the input level. The data sensed by different sensors are combined, and their correlations are analyzed and removed before they are input to a deep learning model. A simple example is the concatenation of sensed data.

Second, decision-level fusion strategies analyze sensed data separately and fuse the individual results at the

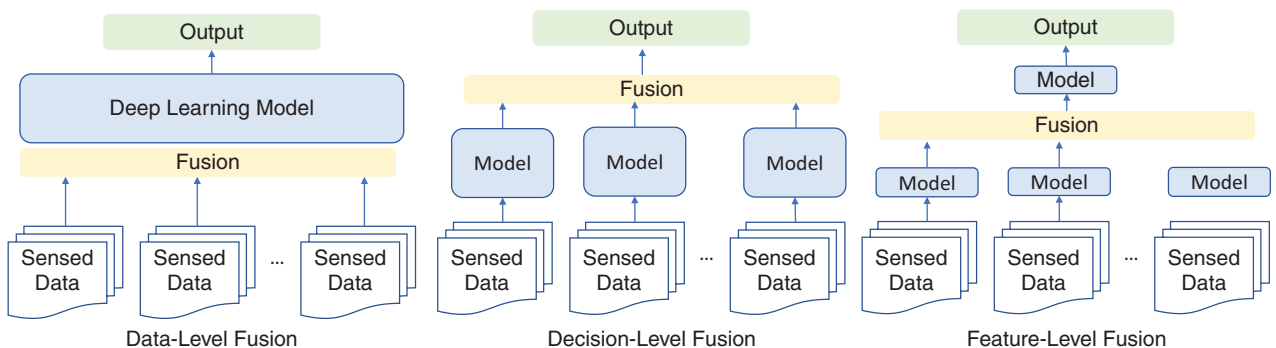


FIGURE 1. The data-level fusion, decision-level fusion, and feature-level fusion strategies.

decision level. An example is the ensemble model, which aggregates results from several deep learning models to accomplish a given task.

Third are feature-level fusion strategies, which offer great flexibility in terms of how data are fused by a deep learning model. The representations of

deep learning has begun. For example, researchers fused MRI and PET images by using a Visual Geometry Group (VGG) network to obtain one resulting image.<sup>2</sup> Information entropy was employed to evaluate the effectiveness of the fused data. In another study, MRI and CT images were

applying deep-learning-based multisensor fusion to human activity recognition, the first challenge is to collect training data for the designated task. Currently, most data sets are proprietary. Second, owing to the limited resources of wearable sensors, the developed models have size restrictions. The model architecture and model size affect the resulting accuracy and response time.

One example of the deployment of deep-learning-based multisensor fusion in personal health monitoring is sleep detection. Researchers have employed deep learning models to fuse EEG and electro-oculogram (EOG) signals to monitor sleep.<sup>5</sup> This monitoring can help with the diagnosis of sleep-related diseases. Mekruksavanich and Jitpattanakul<sup>7</sup> used a convolutional neural network (CNN) to obtain EEG signals as well as a long short-term memory (LSTM) model to obtain EOG signals, and the two signals were fused to generate a final decision. The training data sets included a private data set and the Sleep-European Data Format expanded database,<sup>6</sup> which contains recorded EEG signals, EOG signals, and other sleep physiological signals with manually labeled events.

Another example is the detection of human activities by using various smartphone sensors.<sup>7</sup> In this example, time series data from the accelerometer and gyroscope embedded in a smartphone provide information in the continuous time space, and, for this reason, LSTM-based models are used. The training data set was the University of California, Irvine Human Activity Recognition data set,<sup>8</sup> which classifies six activities on the basis of the sensor data obtained from a waist-mounted smartphone.

### Self-driving applications

With advances in perception sensors, such as cameras as well as radar and lidar sensors, ADASs are supporting diverse functionalities with finer granularity and shorter response times.

Multisensor fusion modules are crucial to ensuring that robots can precisely perceive the physical world and moving objects.

sensed data can be fused across various types of inputs and different layers of a deep learning model. Hence, the model learns how to jointly represent multisensor data. In practice, a mixture of these three types of fusion strategies is feasible. For example, data-level fusion can be combined with feature-level fusion to obtain flexible feature representations within a deep-learning model.

## DEEP-LEARNING-BASED MULTISENSOR FUSION APPLICATIONS

Deep-learning-based multisensor fusion has been investigated in many application fields. Herein, we present a few examples that demonstrate the potential and opportunities created by this fusion technology.

### Medical applications enabled by microelectronic sensing technology

Magnetic resonance imaging (MRI) and computed tomography (CT) are common tools for generating internal images of the human body to enable medical diagnoses. Positron emission tomography (PET) scans can also produce detailed 3D images. MRI relies on radio waves, whereas CT uses X-rays. PET is a type of nuclear medicine procedure and, thus, uses radioactive substances. These are advanced medical technologies for detecting diseases and injuries.

Research on fusing these advanced medical imaging modalities by using

fused to concentrate information from source images into a final image.<sup>3</sup> The fusion of medical images is performed with the aim of improving the usefulness of images when making medical diagnoses. In addition, multisensor fusion has been applied in an end-to-end approach for clinical diagnosis. For instance, T1-weighted imaging, T2-weighted imaging, and fluid-attenuated inversion recovery are MRI scanners with unique features. Researchers fused their features by using deep learning models, with the aim of better classifying glioma,<sup>4</sup> a common type of brain tumor.

### Human activity recognition driven by wearable sensing technology

With the wide availability of wearable devices equipped with diverse sensors, multisensor fusion now plays an important role in human activity recognition within diverse fields,<sup>9</sup> such as personal health monitoring, medical care, and gaming. Data sensed by multiple sensor modalities installed at multiple locations on the human body can be fused to facilitate more comprehensive analyses. Common wearable sensors for human activity recognition include inertial sensors, such as accelerometers and gyroscopes; physiological sensors, such as electroencephalography (EEG) devices; thermal sensors; and pressure sensors. They are often inexpensive, flexible, and easy to integrate by design. When

They overcome blind spots by deploying more sensors with diverse sensing abilities. Object detection and scene segmentation are the two main cornerstones of ADASs. Object detection is used to detect pedestrians, vehicles, and road signs and can be employed to realize various ADAS functions, such as forward-collision avoidance, automatic parking, and traffic sign recognition.

Visual sensor-based object detection can be enhanced by using additional sensors that provide or derive depth information over time, meaning that objects can not only be detected but also be located in 3D space in continuous time. For example, 3D object detection can be achieved using an automobile-mounted stereo camera consisting of two visual cameras,<sup>10</sup> where the proposed stereo region-based CNN model uses feature-level fusion network to fuse images obtained from a stereo camera and predict the 3D bounding boxes of detected objects.

Another approach is to fuse the data streams of a camera and a lidar sensor,<sup>11,12</sup> and this approach is being actively researched with the aim of developing a more robust or more precise object detection model. The Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago data set is widely used in the automotive context.

Scene segmentation tasks help ADASs understand their physical surroundings and contribute to the functions of these systems, such as automatically remaining in and changing lanes. Multisensor fusion can also help manage various weather conditions. For example, by using a deep learning model to fuse the data streams emanating from a visual camera and a thermal sensor, semantic segmentation can be performed in snowy weather.<sup>13</sup> The vision-thermal fusion model detects persons with higher accuracy than a camera-only model in snowy scenarios. This example demonstrates that diverse sensor inputs can be used to compensate for various unsuitable weather conditions in the self-driving context.

## SATELLITE APPLICATIONS ASSISTED BY REMOTE SENSING TECHNOLOGY

Satellite sensors produce images with diverse spatial, spectral, and temporal resolutions owing to the satellites' different orbit altitudes and revisit periods. For example, the Moderate-Resolution Imaging Spectroradiometer has fine granularity in time, whereas Landsat has fine granularity over covered land spaces. More advanced satellite missions offer superior temporal and spatial resolutions and include the Copernicus Sentinel-2 mission. By fusing complementary satellite images, finer Earth observations can be made. Because feature engineering relies heavily on experts' domain knowledge, deep-learning-based

- › First, data collection from multiple sensors is at least as challenging as that from a single sensor. In multisensor fusion, time synchronization among all sensed data is an additional challenge that is not encountered when using a single sensor, especially in the case of complementary sensors. When asynchrony occurs in data sets, it is difficult to detect and recover from.
- › Second, identifying which fusion architecture will offer the highest task performance in advance is challenging. In a relatively popular field, there may be some well-known and

Data sensed by multiple sensor modalities installed at multiple locations on the human body can be fused to facilitate more comprehensive analyses.

satellite fusion opens a door for data scientists who may not have a strong background in satellite technology.

One example of deep-learning-based satellite fusion is superresolution imagery, for which images from Landsat and Sentinel-2 are fused, and time series images with high resolution are generated.<sup>13</sup> Another example is the fusion of satellite images from homogeneous satellites but in different modes for detecting ice-wedge polygons across the Arctic region.<sup>14</sup> Superior Earth observations enable scientists to closely monitor Earth in real time and precisely model planetary changes to generate predictions.

## CHALLENGES


Although massive progress has been made by using deep-learning-based multisensor fusion in various fields, general challenges and challenges specific to each application domain remain. We discuss a few of these challenges and expect to overcome them by conducting more research:

- › well-verified models. However, less research has been conducted on the domain adaptation of multisensor fusion models. In specific scenarios, proprietary studies and massive empirical experiments remain in demand.
- › Cooperative sensors are commonly believed to have higher task performance and robustness against errors or attacks. However, the problem description must be refined to obtain a proper problem statement. For example, some early studies have demonstrated that the use of additional sensors may not lead to the expected level of robustness, and novel, effective attacks against multisensor fusion models have emerged.<sup>16,17</sup> Moreover, multiple sensors can potentially expose a wider range of sensor abnormalities and vulnerabilities. Recovering from faulty or exploited sensors and sensed data is challenging.



**DISCLAIMER**

The author is completely responsible for the content in this message. The opinions expressed here are her own.

In theory, more information leads to more knowledge, which improves the quality of decisions. We reviewed the concept of deep-learning-based multi-sensor fusion and its potential to contribute to various application fields. As sensor technology continues to advance and deep learning technology enables richer multisensor data representation, deep-learning-based multisensor fusion will create new opportunities for increasing efficiency, performance, and robustness. We have highlighted the challenges associated with the development of multisensor fusion in data collection and fusion model design. Furthermore, new forms of vulnerabilities related to multisensor fusion warrant careful investigation. 

**REFERENCES**

1. National Highway Traffic Safety Administration, "Overview of motor vehicle crashes in 2020," U.S. Dept. Transp., Washington, DC, USA, Mar. 2022. [Online]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266>
2. N. Amini and A. Mostaar, "Deep learning approach for fusion of magnetic resonance imaging-positron emission tomography image based on extract image features using pretrained network (VGG19)," *J. Med. Signals Sensors*, vol. 12, no. 1, pp. 25–31, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8804594/>, doi: 10.4103/jmss.JMSS\_80\_20.
3. A. S. Yousif, Z. Omar, U. U. Sheikh, and S. A. Khalid, "A new scheme of medical image fusion using deep convolutional neural network and local energy pixel domain," in *Proc. 2020 IEEE-EMBS Conf. Biomed. Eng. Sci. (IECBES)*, 2021, pp. 384–388, doi: 10.1109/IECBES48179.2021.9398840.
4. G. Chenjie, I. Y.-H. Gu, A. S. Jakola, and J. Yang, "Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks," in *Proc. 2018 IEEE 40th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, pp. 5894–5897, doi: 10.1109/EMBC.2018.8513556.
5. L. Duan *et al.*, "A novel sleep staging network based on data adaptation and multimodal fusion," *Frontiers Hum. Neurosci.*, vol. 15, p. 727139, Oct. 8, 2021, doi: 10.3389/fnhum.2021.727139.
6. B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberyé, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Trans. Bio-Med. Eng.*, vol. 47, no. 9, pp. 1185–1194, 2000, doi: 10.1109/10.867928.
7. S. Mekruksavanich and A. Jitpatanakul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, p. 1636, 2021, doi: 10.3390/s21051636.
8. D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21th Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*, 2013, pp. 437–442.
9. S. Qiu *et al.*, "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges," *Inf. Fusion*, vol. 80, pp. 241–265, Apr. 2022, doi: 10.1016/j.inffus.2021.11.006.
10. P. Li, X. Chen, and S. Shen, "Stereo R-CNN based 3D object detection for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 7636–7644, doi: 10.1109/CVPR.2019.00783.
11. T. Huang, Z. Liu, X. Chen, and X. Bai, "EPNet: Enhancing point features with image semantics for 3D object detection," in *Proc. 2020 16th Eur. Conf. Comput. Vis. (ECCV)*, vol. 12360, pp. 35–52, doi: 10.1007/978-3-030-58555-6\_3.
12. S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 770–779, doi: 10.1109/CVPR.2019.00086.
13. S. Vachmanus, A. A. Ravankar, T. Emaru, and Y. Kobayashi, "Multimodal sensor fusion-based semantic segmentation for snow driving scenarios," *IEEE Sensors J.*, vol. 21, no. 15, pp. 16,839–16,851, 2021, doi: 10.1109/JSEN.2021.3077029.
14. Z. Shao, J. Cai, P. Fu, L. Hua, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ. J.*, vol. 235, p. 111,425, Dec. 2019, doi: 10.1016/j.rse.2019.111425.
15. C. Witharana *et al.*, "Understanding the synergies of deep learning and data fusion of multispectral and panchromatic high resolution commercial satellite imagery for automated ice-wedge polygon detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 170, pp. 174–191, Dec. 2020, doi: 10.1016/j.isprsjprs.2020.10.010.
16. J. Tu *et al.*, "Exploring adversarial robustness of multi-sensor perception systems in self driving," in *Conf. Robot Learn. (CoRL)*, 2021, pp. 1013–1024.
17. J. B. Li, S. Qu, X. Li, P.-Y. Huang, and F. Metze, "On adversarial robustness of large-scale audio visual learning," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 231–235, doi: 10.1109/ICASSP43922.2022.9746124.

**HSIAO-YING LIN** is a principal researcher at Huawei France, Boulogne-Billancourt, 92100, France, and a Member of IEEE. Contact her at [hiaoqing.lin@gmail.com](mailto:hiaoqing.lin@gmail.com).