Sensor fusion with Deep Learning

Jires Donfack Voufo

***Abstract*—**

I. Motivation

Over the years the advancement in transistors technology made it possible to build computers with greater processing power.This also helped to solve more complex issues related to artificial intelligence.AI can be divided in two subsections which are machine learning and deep learning.The decision to use either of this two algorithm or to combine both in a project depend on the type of problem we are trying to solve.Different DL algorithms are getting develop and refine for safety critical applications such as pedestrian detection in vehicles.In this case the DL model has to process and fuse the data from the different sensors to make the appropriate decision related to the type of object it sees.

In this paper we will focus on deep learning with its application in the field of sensor fusion.We will define what sensor fusion is,explore the architecture of a multi-modal sensor fusion network.We will also use a practical example for illustration purpose and give advantages of DL in sensor fusion

II. Foundation

Deep learning extract patterns from data using neural networks with the aim to teach a computer how to learn a task using raw data[1]. The application of DL can be found in image recognition,speech recognition,big data and natural language processing.Autonomous vehicles use sensor fusion with a DL algorithm to be able to identify objects in their surroundings.Sensor fusion consist of combining data coming from multiple sensors.The aim is to get the highest possible accuracy of the parameters being measured or the environment that we are trying to monitor. The major factors dictating the selection of sensors for fusion are: the compatibility of the sensors for deployment within the same environment, and the complementary nature of the information derived from the sensors. If the sensors were to merely duplicate the information, the fusion process would merely be the equivalent of building in redundancy for the enhancement of reliability of the overall system. On the other hand, the sensors should be complementary enough in terms of such variables as data rates, field of view, range, sensitivity to make fusion of information meaningful

III. sensor Fusion

[3] The different types of sensor fusion that exist are:

- Data Fusion:
  This type of sensor fusion combines data from multiple sensors at the raw data level. The goal is to improve the accuracy and precision of the data by combining information from multiple sources.
- Feature Fusion:
  This type of sensor fusion combines data from multiple sensors at the feature level. The goal is to extract relevant features from each sensor and combine them to create a more comprehensive representation of the environment.
- Decision Fusion:
  This type of sensor fusion combines the decisions or conclusions made by multiple sensors. The goal is to improve the overall decision-making process by taking into account multiple sources of information.
- Multi-modal sensor Fusion:
  This type of sensor fusion combines data from multiple sensors that measure different physical phenomena. For example, it can use data from a camera and a lidar sensor to create a more complete representation of the environment.
- Hierarchical sensor Fusion:
  This type of sensor fusion is a process of combining information from multiple sensors at different levels of a system, such as a sensor level, feature level, and decision level.
- Complementary: Here the sensors work in an independent way , but it is still possible to combine their data to get a better image of the environment.
- Competitive: here all sensors measure the same parameter and each of them will give its own measurement.This way we can increase the reliability and accuracy of the system.Due to redundancy in this type of setup,we have to pay attention not get inconsistent readings.
- Cooperative: Sensors are cooperative when they provide independent measurements, that when combined provide information that would not be available from any one sensor. Cooperative sensor networks take data from simple sensors and construct a new abstract sensor with data that does not resemble the readings from any one sensor.

A. *early fusion*

Early fusion is applicable on raw data or pre-processed data obtained from sensors. Data features should be extracted from the data before fusion, otherwise the process will be challenging especially when the data sources have different sampling rates between the modalities. Synchronization of data sources is also challenging when one data source is discrete and the others are continuous. Hence, converting data sources

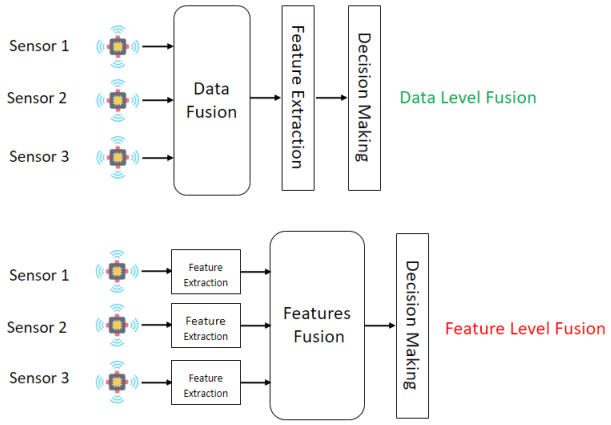

Fig. 1.

into a single feature vector is a significant challenge in early data fusion.

B. intermediate fusion

C. late fusion

Late fusion uses data sources independently followed by fusion at a decision-making stage (Figure 4). Late data fusion is inspired by the popularity of ensemble classifiers [11]. This technique is much simpler than the early fusion method, particularly when the data sources are significantly varied from each other in terms of sampling rate, data dimensionality and unit of measurement. Late fusion often gives better performance because errors from multiple models are dealt with independently

IV. Neural network

A. Structure

B. Convolutional Neural network

A Convolutional Neural Network (CNN) is generally used for input data that has a natural grid-like pattern. The most common input data for a CNN is images. One feature that distinguishes a CNN from a regular network is that it can extract features The three of layers of CNN are as follow

- convolution layer and pooling layer: for feature extraction
- convolution layer and pooling layer: can possess one or more layers to map features to the output values

The first layer of any network is the input layer. In the case of a CNN, the input layer is commonly an image which has been pre-processed to fit the architecture's expected values .The next layer is the convolutional layer, consisting of linear operations for feature extraction. The idea is to use a kernel represented as a small 2D matrix, usually of odd dimension. This kernel is used to obtain the relationship between the center pixel of a box with its neighboring pixels. The convolutional layer slides the kernel across the larger input image with the goal of modifying the middle pixel of the larger input image .The number of strides determines the amount to shift the kernel, and padding refers to the technique of adding zeros to enable convolution on each pixel. Both of these operations affect the size of the output of the convolutional layer based on the configuration.

C. Object recognition

Here a residual Neural network is usualy used for deep feature extraction.This network makes it posssible to skip any layer that is having a negative impact in the training process

The pooling layer follows the convolutional layer to downsize or downsample the feature map into a smaller matrix. This is done by pooling the values in adjacent pixels [29]. Max pooling is the more popular technique used in research [28]. This is followed by a non-linear activation, commonly a ReLU function, to ensure the feature map contains no negative values [29]. The process of convolution, pooling, and activation is repeated until a small matrix is obtained such that the activation is highly dependent on more complex features. The final layer is the fully connected layer. The input to this layer is the small matrix obtained from the previous layers. This matrix is then flattened, i.e., converted to a 1D array. The fully connected layer, as stated in its name, is designed such that every input is connected to every output by a weight. The number of output neurons is determined by the number of classes [29]. The activation function for this final layer is generally the softmax function for multi-class classification.

D. Neural Network training

Training a neural network means that we start with a bad performing network and as the model learns from the data and ground truth values, we update the network such that we end with a highly accurate model. The loss or error rate is used to update the weights of the model using backpropagation as described above. Optimizers like the Stochastic Gradient Descent (SGD) or Adam that use gradient-based algorithms are used to minimize the loss, which is equivalent to maximizing the performance

E. Neural Network Evaluation

Here we have to evaluate how many correct prediction the train model makes

The architecture of the multi-modal network is as follow:[2]

- Uni-modal neural network to process the different sensors input separately.
- Fusion network(combines the different features extracted)
- Classifier network(use the classified data to make decisions)

V. Practical case

In this use case We consider some objects that are to be collected by a robot.The decision of the order in which the objects are to be picked is made by the robot with the help of sensor fusion and deep learning.The 3 sensors to be use by the robot are: 2 cameras, accelerometer and encoder.

A. *early sensor fusion*

this phase is divided in 3 steps:

- Combination of data:the distance as calculated by cameras,accelerometer and encoders are put together while taking in account the weigh of each sensor
- Extraction:the features of collected data get extracted using 1D convolution
- Classification:(use the classified data to make decisions)

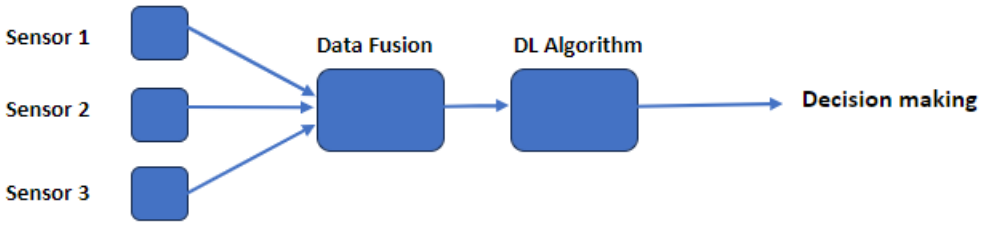

Fig. 2.

VI. Advantages Of using DL in Sensor fusion

Systems that use sensor fusion with DL are more consistent and accurate.

VII. Conclusion

References

References

[1] https://de.mathworks.com/videos/sensor-fusion-part-1-what-is-sensor-fusion-1569410785813.html

[2]