AGENT NETWORK COMPRISING ONE OR MORE AGENTS AND COORDINATOR AND METHOD FOR AGENT NETWORK

TECHNICAL FIELD

The present disclosure relates generally to the field of wireless communication and more specifically, to an agent network comprising one or more agents and a coordinator. Furthermore, the present disclosure relates more specifically to a method for the agent network comprising one or more agents and the coordinator.

BACKGROUND

Advancements in wireless communications have played a critical role in shaping the capabilities of agents, especially in scenarios where real-time decision-making is crucial, especially in autonomous vehicles and robotics. Autonomous agents of the autonomous vehicles are required to perform certain actions (e.g., actions required for safe navigation, industrial assembling, and the like) based on distributed sensing data collected from sensors in real-time. However, optimal actions require full knowledge about an environment that includes the intentions of other active autonomous agents, for example, if another autonomous vehicle wants to change a lane. The autonomous agents have partial knowledge about the environment but do not have any information about the other autonomous agents, which increases the security risks.

Conventionally, certain attempts have been made to enable the autonomous vehicles to collaborate effectively in dynamic environments. However, such attempts fail due to many reasons, such as decentralized processing of locally collected data, incapability of collecting data from moving sensors in real-time, support of heterogeneous data formats and data types, limitation of resources, and the like. As a result, there exists a technical problem of how to provide an improved collaboration and decision-making abilities of the autonomous agents in real-time environments while adapting to dynamic conditions.

Therefore, in light of the foregoing discussion, there exists a need to overcome the aforementioned drawbacks associated with the conventional agent network and conventional methods.

SUMMARY

The present disclosure provides an agent network comprising a one or more agents and a coordinator. Furthermore, the present disclosure relates more specifically to a method for the agent network comprising the one or more agents and the coordinator. The present disclosure provides a solution to the existing problem of how to provide an improved collaboration and decision-making abilities of the autonomous agents in real-time environments while adapting to dynamic conditions. An objective of the present disclosure is to provide a solution that overcomes at least partially the problems encountered in the prior art and provides an improved agent network and an improved method for the agent network, such as by generating a global scene model that can be further utilized for an accurate and reliable decision making.

One or more objectives of the present disclosure are achieved by the solutions provided in the enclosed independent claims. Advantageous implementations of the present disclosure are further defined in the dependent claims.

In one aspect, the present disclosure provides an agent network including one or more agents and a coordinator. Moreover, each agent includes a sensor and a controller. The controller is configured to receive sensor input from the sensor of a partial scene, identify features relevant to solving an action, generate a segmented data set and a low-resolution sensor data set, and send the segmented data set and the low-resolution sensor data set to the coordinator. The coordinator includes a controller configured to receive a first segmented data set and a first low-resolution sensor data set from a first agent. Furthermore, the controller is configured to receive a second segmented data set and a second low-resolution sensor data set from a second agent and generate a model of a global scene based on a combination of the received data sets from the agents. The controller is configured to receive a query from the first agent, determine a prompt for the query based on the global scene model, and transmit the prompt to the first agent. Moreover, the controller of the first agent is further configured to receive the prompt, reevaluate the sensor data in the view of the received prompt, and perform a desired action.

Advantageously, the agent network is configured to generate a comprehensive global scene model of the dynamic environment based on the diverse local views of the autonomous agents. Furthermore, utilize the generated global scene model to enhance the decision-making abilities of the one or more agents. Each agent is equipped with sensors and controllers that are configured to capture the image data, which can be further analysed to perform an action. Moreover, the agent network includes the coordinator that is configured to combine the data sets (i.e., the segmented data set and the low-resolution sensor data set) from the one or more agents of the agent network in order to generate the global scene model. The generation of the global scene model enables the one or more agents of the agent network to take informed decisions based on a collective understanding of the agent environment, leading to efficient, effective, and coordinated actions. Moreover, the one or more agents of the agent network are free to join or leave the agent network based on the requirement due to which the agent network offers high flexibility. The one or more agents in the agent network contribute to a high bandwidth optimization by transmitting the data to the other one or more agents and the coordinator, when required in order to minimize the amount of data exchanged within the agent network and enhance the optimized resource utilization. Furthermore, the one or more agents enhance the privacy of the agent network by selectively sending the segmented data and the low-resolution sensor data set to the coordinator of the one or more agents instead of directly sharing the raw data with the coordinator of one or more agents. Moreover, the determination of the prompt for the query based on the global scene model allows the one or more agents of the agent network to seek additional information that is further utilized to perform the desired action efficiently and reliably. As a result, the agent network is configured to provide a scalable, resource-efficient, and reliable solution for an accurate and efficient decision-making by the one or more agents of the agent network.

In another aspect, the present disclosure provides a method for an agent network comprising one or more agents and a coordinator and each agent comprises a sensor. Moreover, the method comprises the agent for receiving sensor input from the sensor of a partial scene, identifying features relevant for solving an action, generating a segmented data set and a low-resolution sensor data set, and sending the segmented data set and the low-resolution sensor data set to the coordinator. Furthermore, the method comprises the coordinator for receiving a first segmented data set and a first low-resolution sensor data set from a first agent, receiving a second segmented data set and a second low-resolution sensor data set from a second agent, generating a model of a global scene based on a combination of the received data sets from the agents, and receiving a query from the first agent, determining a prompt for the query based on the model of a global scene. After that, transmitting the prompt to the first agent, whereby the method further comprises the first

agent for receiving the prompt, reevaluating the sensor data in the view of received prompt, and performing a desired action.

The disclosed method achieves all the advantages and technical effects of an agent network comprising one or more agents and a coordinator.

In yet another aspect, the present disclosure provides an agent in an agent network comprising one or more agents and a coordinator. Moreover, each agent comprises a sensor and a controller configured to receive sensor input from the sensor of a partial scene, identify features relevant for solving an action, generate a segmented data set and a low-resolution sensor data set and send the segmented data set and the low-resolution sensor data set to the coordinator. Furthermore, the agent is configured to send a query to the coordinator, receive a prompt based on a global scene for the query, reevaluate the sensor data in the view of received prompt and perform a desired action.

Advantageously, the agent is configured to provide a dynamic query-prompt interaction with the coordinator, allowing for real-time adaptation of the agent's decision-making based on the received prompt, thereby enhancing its responsiveness and adaptability in complex scenarios within the agent network.

In another aspect, the present disclosure provides a method for an agent in the agent network comprising one or more agents and a coordinator. Moreover, each agent includes a sensor and the method includes an agent for receiving sensor input from the sensor of a partial scene, identifying features relevant for solving an action, generating a segmented data set and a low-resolution sensor data set and sending the segmented data set and the low-resolution sensor data set to the coordinator, sending a query to the coordinator, receiving a prompt based on a global scene for the query, reevaluating the sensor data in the view of received prompt, and performing a desired action.

The disclosed method achieves all the advantages and technical effects of an agent of the agent network comprising one or more agents and a coordinator.

In yet another aspect, the present disclosure provides a coordinator in an agent network comprising one or more agents and each agent comprises a sensor. Furthermore, the coordinator comprises a controller configured to receive a first segmented data set and a first low-resolution sensor data set from a first agent, receive a second segmented data set and a second low-resolution sensor data set from a second agent, generate a model of a global scene based on a combination of the received data sets from the agents, receive a query from the first agent, determine a prompt for the query based on the global scene model, and transmit the prompt to the first agent.

Advantageously, the coordinator is configured to generate a comprehensive model of the global scene by combining data from multiple agents, enabling prompt-based interaction and tailored guidance to individual agents within the network, enhancing overall decision-making capabilities.

In yet another aspect, the present disclosure provides a method for coordinator in an agent network comprising one or more agents and each agent comprises a sensor, wherein the method comprises the coordinator receiving a first segmented data set and a first low-resolution sensor data set from a first agent, receiving a second segmented data set and a second low-resolution sensor data set from a second agent, generating a model of a global scene based on a

combination of the received data sets from the agents, receiving a query from the first agent, determining a prompt for the query based on the global scene model, and transmitting the prompt to the first agent.

The disclosed method achieves all the advantages and technical effects of a coordinator of the agent network comprising one or more agents and a coordinator.

It is to be appreciated that all the aforementioned implementation forms can be combined.

It has to be noted that all devices, elements, circuitry, units, and means described in the present application could be implemented in the software or hardware elements or any kind of combination thereof. All steps which are performed by the various entities described in the present application as well as the functionalities described to be performed by the various entities, are intended to mean that the respective entity is adapted to or configured to perform the respective steps and functionalities. Even if, in the following description of specific embodiments, a specific functionality or step to be performed by external entities is not reflected in the description of a specific detailed element of that entity which performs that specific step or functionality, it should be clear for a skilled person that these methods and functionalities can be implemented in respective software or hardware elements, or any kind of combination thereof. It will be appreciated that features of the present disclosure are susceptible to being combined in various combinations without departing from the scope of the present disclosure as defined by the appended claims.

Additional aspects, advantages, features, and objects of the present disclosure would be made apparent from the drawings and the detailed description of the illustrative implementations construed in conjunction with the appended claims that follow.

BRIEF DESCRIPTION OF THE DRAWINGS

The summary above, as well as the following detailed description of illustrative embodiments, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the present disclosure, exemplary constructions of the disclosure are shown in the drawings. However, the present disclosure is not limited to specific methods and instrumentalities disclosed herein. Moreover, those in the art will understand that the drawings are not to scale. Wherever possible, like elements have been indicated by identical numbers.

Embodiments of the present disclosure will now be described, by way of example only, with reference to the following diagrams wherein:

FIG. 1 is a block diagram that depicts an agent network comprising one or more agents and a coordinator, in accordance with an embodiment of the present disclosure;

FIGs. 2A and 2B are diagrams that depict a flow chart of a method for an agent network comprising one or more agents and a coordinator, in accordance with an embodiment of the present disclosure;

FIG. 3 is a block diagram that depicts an agent in an agent network comprising one or more agents, in accordance with an embodiment of the present disclosure;

FIG. 4 is a diagram that depicts a flow chart of a method for an agent in an agent network comprising one or more agents, in accordance with an embodiment of the present disclosure;

FIG. 5 is a block diagram that depicts a coordinator in the agent network, in accordance with an embodiment of the present disclosure;

FIG. 6 is a diagram that depicts a flow chart of a method for a coordinator in the agent network, in accordance with an embodiment of the present disclosure;

FIG. 7A is a diagram depicting a recreation of the global scene model, in accordance with an embodiment of the present disclosure;

FIG. 7B is a diagram depicting a local view processed by an agent, in accordance with an embodiment of the present disclosure;

FIG. 8 is a diagram depicting an input processing of the agent, in accordance with an embodiment of the present disclosure;

FIG. 9 is a diagram that illustrates global scene model regeneration using cross-view attention and diffusion, in accordance with an embodiment of the present disclosure;

FIG. 10 is a diagram that illustrates a prompt generation by a coordinator, in accordance with an embodiment of the present disclosure; and

FIG. 11 is a diagram that illustrates a scene-aware semantic analysis at the end of an agent, in accordance with an embodiment of the present disclosure.

In the accompanying drawings, an underlined number is employed to represent an item over which the underlined number is positioned or an item to which the underlined number is adjacent. A non-underlined number relates to an item identified by a line linking the non-underlined number to the item. When a number is non-underlined and accompanied by an associated arrow, the non-underlined number is used to identify a general item at which the arrow is pointing.

DETAILED DESCRIPTION OF EMBODIMENTS

The following detailed description illustrates embodiments of the present disclosure and ways in which they can be implemented. Although some modes of carrying out the present disclosure have been disclosed, those skilled in the art would recognize that other embodiments for carrying out or practicing the present disclosure are also possible.

FIG. 1 is a block diagram that depicts an agent network comprising one or more agents and a coordinator, in accordance with an embodiment of the present disclosure. With reference to FIG. 1A, there is shown an agent network **100** that includes one or more agents **102**, a coordinator **104**, and a communication network **106**.

The one or more agents **102** refers to a pool of autonomous agents that are capable of sensing the dynamic environment, process information, and take the required necessary actions. In an implementation, the one or more agents **102** are autonomous agents of autonomous vehicles. Moreover, the one or more agents **102** includes a first agent **102A**, and a second agent **102B** up to Nth agent configured to collect and process data. Each agent from the one or more agents **102** further includes a sensor and a controller. For example, the first agent **102A** includes a first controller **108**, a first sensor **110**, and a first memory **112**. Similarly, the second agent **102B** includes a second controller **118** a second sensor **120**, and a second memory **122**.

The coordinator **104** is configured to generate a global scene model **132** based on a combination of the received data sets from the agents. The coordinator **104** includes a third controller **128**, a third memory **130** to store the global scene model **132**, and a network interface **134**.

The communication network **106** includes a medium (e.g., a communication channel) through which the one or more agents **102** and the coordinator **104** communicate with each other. Examples of the communication network **106** may include, but are not limited to, a cellular network (e.g., a 2G, a 3G, long-term evolution (LTE) 4G, a 5G, or 5G New Radio (NR) network, such as sub 6 GHz, cmWave, or mmWave communication network), a wireless sensor network (WSN), a cloud network, a Local Area Network (LAN), a vehicle-to-network (V2N) network, a Metropolitan Area Network (MAN), and/or the Internet.

The first controller **108** of the first agent **102A** is configured to receive a sensor input from the first sensor **110** of a partial scene. Similarly, the second controller **118** of the second agent **102B** is configured to receive a sensor input from the second sensor **120** of a partial scene. Furthermore, the third controller **128** of the coordinator **104** is configured to generate the global scene model **132** based on the combination of the received dataset from the one or more agents **102**. Examples of the first controller **108**, the second controller **118**, and the third controller **128** may include but are not limited to a central data processing device, a microprocessor, a microcontroller, a complex instruction set computing (CISC) processor, an application-specific integrated circuit (ASIC) processor, a reduced instruction set (RISC) processor, a very long instruction word (VLIW) processor, a central processing unit (CPU), a state machine, a data processing unit, and other processors or circuitry.

The first memory **112** of the first agent **102A** is configured to store a first segmented data set **114** and a first low-resolution sensor data set **116**. Similarly, the second memory **122** of the second agent **102B** is configured to store a second segmented data set **124** and a second low-resolution sensor data set **126**. Furthermore, the third memory **130** is configured to store the global scene model **132**. Examples of implementation of the first memory **112**, the second memory **122**, and the third memory **130** may include, but are not limited to, Electrically Erasable Programmable Read-Only Memory (EEPROM), Dynamic Random-Access Memory (DRAM), Random Access Memory (RAM), Read-Only Memory (ROM), Hard Disk Drive (HDD), Flash memory, a Secure Digital (SD) card, Solid-State Drive (SSD), and/or CPU cache memory.

The network interface **134** may include hardware or software that is configured to establish communication between the third controller **128** and the third memory **130**. Examples of the network interface **134** may include but are not limited to a computer port, a network socket, a network interface controller (NIC), and any other network interface device.

There is provided the agent network **100** that includes the one or more agents **102** and the coordinator **104**. Moreover, each agent from the one or more agents **102** includes a sensor and a controller configured to receive sensor input of a partial scene, identify features relevant for solving an action, generate a segmented data set and a low-resolution sensor data set, and send the segmented data set and the low-resolution sensor data set to the coordinator **104**. In an example, the first agent **102A** includes the first sensor **110** and the first controller **108** that is configured to receive a sensor input from the first sensor **110** of the partial scene. Thereafter, the first controller **108** is configured to identify features relevant for solving the action and generate the first segmented data set **114** and the first low-resolution sensor data set **116**. Furthermore, the first controller **108** is configured to send the first segmented data set **114** and the first low-resolution sensor data set **116** to the coordinator **104**. In another example, the second agent **102B** includes the second sensor **120** and the second controller **118** that is configured to receive the sensor input from the second sensor **120** of the partial scene. Thereafter, the second controller **118** is configured to identify features relevant for solving an action

and generate the second segmented data set **124** and the second low-resolution sensor data set **126**. Furthermore, the second controller **118** is configured to send the second segmented data set **124** and the second low-resolution sensor data set **126** to the coordinator **104**. Moreover, the received sensor input from each of the one or more agents **102** is used to provide a comprehensive and detailed understanding of the dynamic environment in real-time that can be further utilized for recognizing objects, patterns, or any other relevant elements in the dynamic environment. Furthermore, the generation of the segmented data set (e.g., the first segmented data set **114** and the second segmented data set **124**) and low-resolution sensor data set (i.e., the first low-resolution sensor data set **116** and the second low-resolution sensor data set **126**) are used to get an insight on a specific aspect of the partial scene of the dynamic environment. Furthermore, the generated segmented data set and the low-resolution sensor data set are transmitted to the coordinator **104** to provide a holistic representation of the dynamic environment. In accordance with an embodiment, the sensor is a RADAR sensor, and the sensor data is RADAR image data. In an implementation, the RADAR sensor is configured to use radio waves in order to detect and determine the distance, speed, direction, and other characteristics of objects in the vicinity of the agent. For example, the first sensor **110** of the first agent **102A** is a RADAR sensor and the sensor data is RADAR image data. Similarly, the second sensor **120** is another RADAR sensor and the sensor data is another RADAR image data. Advantageously, the RADAR sensor of each of the one or more agents **102** is used to provide improved object detection and object recognition in adverse weather conditions, such as rain or fog, which can be challenging for other types of sensors. In accordance with another embodiment, the sensor is an image sensor, and the sensor data is image data. In an implementation, the first sensor **110** of the first agent **102A** and the second sensor **120** of the second agent **102B** are configured to take a snapshot of a corresponding partial scene (or local view). Moreover, the first controller **108** of the first agent **102A** and the second controller **118** of the second agent **102B** are configured to collect the snapshots taken by the first sensor **110** and the second sensor **120** respectively. After that, the first controller **108** and the second controller **118** are configured to execute a task-aware processing algorithm that highlights the information related to the actions that are required to be taken by the first agent **102A** and the second agent **102B** while suppressing irrelevant details. Moreover, the task-aware processing algorithm is configured to produce the segmented image, for example, Im-Seg, illustrating detected objects relevant to the task (e.g., cars, pedestrians, street lines, and the like.).

In accordance with an embodiment, the controller (i.e., the first controller **108**) of the first agent **102A** is further configured to determine that the sensor input is insufficient to solve the action and, in response, perform full segmentation using a trained neural network. Furthermore, the first controller **108** is configured to filter out segmented objects unrelated to a local objective of the task and transmit positional information about the agent's global position and orientation, and a timestamp indicating a moment of recording to the coordinator **104**. In an implementation, the first controller **108** of the first agent **102A** is configured to employ a trained neural network, such as Mask R-CNN neural network, and like) to perform a comprehensive segmentation of the input image data, which includes filtration of objects that are irrelevant to the local objective and the relevant objects that are identified as relevant. The goal-aware filtration is further used to generate the first segmented data set **114** and the second segmented data set **124**. Moreover, the input image data undergoes compression using a task-agnostic image compression technique to generate the first low-resolution sensor data set **116** and the second low-resolution sensor data set **126**. As a result, the first segmented data set **114** and the first low-resolution sensor data set **116** are generated to provide information about the global position of the first agent **102A** and its orientation, along with a timestamp indicating the moment when the corresponding image was captured. In addition, the first controller **108** and the second controller **118** are configured to perform the task-agnostic compression algorithm that eliminates unnecessary information from the sensing data,

which generates a low-resolution version of the input image data, for example, Im-LowRes, and the like. By virtue of performing the task-aware processing algorithm and the task-agnostic compression algorithm for the input image data set from the sensor (i.e., the first sensor **110** and the second sensor **120**) enhances the positional embedding and orientation of the first agent **102A**. Furthermore, the inclusion of a timestamp indicates the specific moment of image capturing, contributing to a comprehensive and context-rich dataset for improved decision-making by the first agent **102A** from the one or more agents **102**.

Furthermore, the coordinator **104** includes a controller (i.e., the third controller **128**), which is configured to receive the first segmented data set **114** and the first low-resolution sensor data set **116** from the first agent **102A**. Thereafter, receive the second segmented data set **124** and the second low-resolution sensor data set **126** from the second agent. Furthermore, after receiving the first segmented data set **114**, the first low-resolution sensor data set **116**, the second segmented data set **124**, and the second low-resolution sensor data set **126** the third controller **128** of the coordinator **104** is configured to generate a global scene model **132** based on a combination of the received data sets from the agents (i.e., from the first agent **102A** and the second agent **102B**). For example, the third controller **128** of the coordinator **104** is configured to receive the processed image data from the agent and combine the processed data with information received from other agents. The arrangement of the combined image data is organized using received meta-information such as agent-id, timestamp, and location-agent. Furthermore, the combined data are processed consecutively by two or more algorithms, such as a cross-view attention neural network, a diffusion attention neural network, and like without affecting the scope of the present disclosure.

In accordance with an embodiment, the controller of the coordinator (i.e., the third controller **128**) is further configured to generate the global scene model **132** based on the combination of the received data sets from the agents by running cross-view attention and diffusion. For example, the third controller **128** is configured to generate the global scene model **132** based on the combination of the received data sets, such as the first segmented data set **114**, the first low-resolution sensor data set **116**, the second segmented data set **124**, and the second low-resolution sensor data set **126** that is received from the first agent **102A** and the second agent **102B** by running the cross-view attention and diffusion. The cross-view attention enables the association of the received data sets from the one or more agents **102** to provide a detailed and comprehensive and detailed understanding of the dynamic environment. Similarly, the diffusion is also used to generate the global scene model **132** by filling potential information gaps in the data. As a result, the generation of the global scene model **132** based on the combination of the received data sets from the one or more agents **102** by running the cross-view attention and diffusion is used to ensure a robust and holistic representation of the dynamic environment that is used to enhance the decision-making capabilities of the agent network **100** by offering a more nuanced and accurate global scene model **132**.

Furthermore, the controller of the coordinator (i.e., the third controller **128**) is further configured to receive a query from the first agent **102A**, determine a prompt for the query based on the global scene model **132**, and transmit the prompt to the first agent **102A**. Firstly, the first agent **102A** is configured to collect the local data and further evaluate if the local data has enough information to perform a desired action or not. Moreover, if the local information is insufficient, then, in that case, the first agent **102A** is configured to select a query (i.e., the query from a pre-defined query-list) that describes the required missing information. In an example, the received query form, the first agent **102A** includes "Can I safely turn right?", "Are there some obstacles ahead?", and the like. Thereafter, the first agent **102A** is configured to send the query that includes a tuple having an agent-id, a timestamp, a location-agent, and a

query-id to the coordinator **104**. In an implementation, the query includes information about objects, attributes and relations for objects, topographical information, and additional background information. Moreover, the query is accompanied by locally processed data (e.g., Im-Seg, Im-LowRes, and the like). After that, the coordinator **104**, upon receiving the query and the local data from the first agent **102A**, is configured to extract all the information from the global scene model **132** (or a global database) that is related to the query and the received local data, such as information about objects, their attributes and relations, some topographical information of streets and buildings, and additional background information, for example, bad weather conditions by using data-retrieval techniques. Furthermore, the extracted information forms a list of database records called prompt. In an example, the format of such database records includes obj-id, obj-location, obj-type, and obj-attributes where obj is an object that is described in the global scene model **132**. Finally, the coordinator **104** is configured to send back the prompt to the first agent **102A**. Moreover, the prompt includes an id-agent, a timestamp, a query-id, and a prompt. The first agent **102A** upon receiving the prompt is configured to re-evaluate its own objective by considering newly received information. As a result, the coordinator **104** is configured to retrieve the relevant information from the global scene model **132** and further organize the received information into a prompt, which is shared with the first agent **102A**.In accordance with an embodiment, the controller of the first agent (i.e., the first controller **108**) is further configured to determine that the sensor input is insufficient to solve the action and, in response, accordingly, sends the query to the coordinator **104**. Firstly, the first controller **108** of the first agent **102A** is configured to determine if the available sensor input is insufficient to successfully execute the required action or not. Thereafter, the first controller **108** is configured to generate a query and further transmit the generated query to the coordinator **104**. As a result, the first controller **108** is configured to ensure proactive communication from the first agent **102A** even with data limitations, seeking assistance from the coordinator **104** to enhance the understanding and decision-making capabilities of the first agent **102A**. By empowering the first agent **102A** to recognize the limitations of the sensor input and request assistance through queries to the coordinator **104**, the agent network **100** is configured to promote efficient problem-solving and decision-making in a real-time dynamic environment.

In accordance with an embodiment, the query-related data include one, some, or all of the information about objects, attributes and relations for objects, topographical information, and additional background information. The inclusion of information concerning objects, attributes, relations for objects, topographical details, and additional background information in response to queries helps the first agent **102A** to obtain a comprehensive and contextually rich data when seeking information from the coordinator **104**. The incorporation of diverse query-related data ensures that the first agent **102A** receives a holistic understanding of the corresponding query, such as by including object details, attributes of the objects, relations, topographical features, and additional contextual background. As a result, by including the data, the first agent **102A** is configured to ensure that the received sensor data input includes comprehensive insights that enhance the adaptability of the agent network **100** and further enhance the decision-making capabilities of the first agent **102A**, allowing the first agent **102A** to make informed choices in a complex and dynamic environment.

In accordance with an embodiment, the controller of the coordinator (i.e., the third controller **128**) is further configured to determine the prompt for the query based on the global scene model **132** by extracting query-related data from the global scene model **132**. Moreover, the extracted data is then structured into a prompt tailored to the specific query, ensuring that the first agent **102A** receives precise and pertinent information to guide the decision-making ability of the first agent **102A**. Therefore, by extracting the query-related data directly from the global scene model **132**, the

coordinator **104** is configured to provide accurate and reliable prompts with an enhanced adaptability and decision-making efficiency of the one or more agents **102** in the agent network **100** in the dynamic environment.

In accordance with an embodiment, the controller of the coordinator (i.e., the third controller **128**) is further configured to extract query-related data using data-retrieval techniques. The utilization of the data-retrieval techniques (e.g., semantic data association, pattern recognition, or any other sophisticated algorithms) are used to pinpoint and extract query-related data from the global scene model **132**. The extracted data are then structured into a comprehensive prompt that is transmitted back to the querying agent. As a result, the coordinator **104** is configured to enhance the precision of the extracted data and improve the decision-making efficiency of the agents throughout the agent network **100**.

Furthermore, the controller of the first agent (i.e., the first controller **108**) is further configured to receive the prompt, reevaluate the sensor data in the view of received prompt, and then perform a desired action. Moreover, after receiving the prompt, the first agent **102A** is configured to exploit the newly received information to reevaluate the input data and perform the desired action. By re-evaluating the sensor data in response to specific guidance, the first agent **102A** is configured to improve the responsiveness and decision-making ability of the first agent **102A** within the agent network **100** enhancing the overall adaptability and intelligence of the agent network **100** in navigating complex and dynamic environment.

In accordance with an embodiment, the controller (i.e., the first controller **108**) of the first agent **102A** is further configured to reevaluate the sensor data in the view of the received prompt by assessing the transformed prompt combined with the identified features jointly using a neural network providing an enriched set of features and then selecting the desired action based on the enriched set of features. The received prompt is transformed (i.e., vectorized) to a format, which is acceptable by the agent network **100**. Moreover, the transformed prompt is further combined with the input features and jointly assessed features, such as by using the neural network to select the desired action. The neural network facilitates the combination of the prompt and the identified features with the identified features jointly using a neural network providing an enriched set of features. The first controller **108** of the first agent **102A** is further configured to select the desired action based on the enriched set of features and ensure an accurate and reliable selection of the desired action.

In accordance with an embodiment, the controller of the first agent (i.e., the first controller **108**) is further configured to identify features through image segmentation. The image segmentation refers to a process that partitions an image into distinct segments, allowing the identification of individual elements or objects within the image. The utilization of the image segmentation allows the first controller **108** of the first agent **102A** to delineate and recognize specific features, enhancing the perception of the dynamic environment that becomes valuable input for subsequent decision-making processes within the agent network **100**.

Advantageously, the agent network **100** is configured to generate a comprehensive global scene model **132** of the dynamic environment based on the diverse local views of the autonomous agents. Furthermore, utilize the generated global scene model **132** to enhance the decision-making abilities of the one or more agents **102**. Each agent is equipped with sensors and controllers that are configured to capture the image data, which can be further analysed to perform an action. Moreover, the agent network **100** includes the coordinator **104**, which is configured to combine the data

sets (i.e., the segmented data set and the low-resolution sensor data set) from the one or more agents **102** of the agent network **100** to generate the global scene model **132**. The generation of the global scene model **132** enables the one or more agents **102** of the agent network **100** to take informed decisions based on a collective understanding of the agent environment, leading to an efficient, effective, and coordinated actions. Moreover, the one or more agents **102** of the agent network **100** are free to join or leave the agent network **100** based on the requirement due to which the agent network **100** offers high flexibility. The one or more agents **102** in the agent network **100** contribute to a high bandwidth optimization by transmitting the data to the other one or more agents **102** and the coordinator **104**, when required in order to minimize the amount of data exchanged within the agent network **100** and enhances the optimized resource utilization. Furthermore, the one or more agents **102** enhances the privacy of the agent network **100** by selectively sending the segmented data and the low-resolution sensor data set to the coordinator **104** of the one or more agents **102** instead of directly sharing the raw data with the coordinator **104** of the one or more agents **102**. Moreover, the determination of the prompt for the query based on the global scene model **132** allows the one or more agents **102** of the agent network **100** to seek additional information that is further utilized to perform the desired action efficiently and reliably. As a result, the agent network **100** is configured to provide a scalable, resource-efficient, and reliable solution for accurate and efficient decision-making by the one or more agents **102** of the agent network **100**.

FIG. 2A and FIG. 2B are diagrams that depict a flow chart for the method of an agent network comprising one or more agents and a coordinator. With reference to FIG. 2, there is shown a flowchart of method **200** of an agent network comprising the one or more agents **102** and the coordinator **104**. The method **200** includes steps **202** to **226**. In an implementation, the first controller **108** is configured to execute all the operations of the method **200**.

At step **202**, the method **200** includes receiving the sensor input from the sensor of a partial scene. Moreover, the first agent **102A** and the second agent **102B** are configured to receive the sensor input from the first sensor **110** and the second sensor **120**. Furthermore, the received sensor input of the partial scene may consist of the local image data captured by the first sensor **110** and the second sensor **120**. At step **204**, the method **200** includes, identifying features relevant for solving an action. Moreover, after receiving the sensor input of the partial scene from the sensor, the first agent **102A** and the second agent **102B** are configured to identify the relevant features from the partial scene captured by the first sensor **110** and the second sensor **120**. At step **206**, the method **200** includes, generating the segmented data set and the low-resolution sensor data set. Moreover, the first controller **108** and the second controller **120**of the first agent **102A** and the second agent **102B** are configured to generate the first segmented data set **114**, the first low-resolution sensor data set **116** and the second segmented data set **124**, the second low-resolution sensor data set **126**. At step **208**, the method **200** includes, sending the segmented data set and the low-resolution sensor data set to the coordinator. Moreover, the first agent **102A** and the second agent **102B** are configured to send the first segmented data set **114**, the first low-resolution sensor data set **116** and the second segmented data set **124**, the second low-resolution sensor data set **126** to the coordinator **104**. At step **210**, the method **200** includes, receiving the first segmented data set **114** and the first low-resolution sensor data set **116** set from the first agent **102A**. Moreover, the coordinator **104** in the agent network is configured to receive the first segmented data set **114**, the first low-resolution sensor data set **116** from the first agent **102A**.

At step **212**, the method **200** includes, receiving the second segmented data set and the second low-resolution sensor data set from the second agent. Moreover, the coordinator **104** in the agent network is configured to receive the second segmented data set **124** and the second low-resolution sensor data set **126** from the second agent **102B**. At step **214**,

the method **200** includes, generating a model of a global scene based on the combination of received data sets from agents. Moreover, the coordinator **104** is configured to generate a model of the global scene **132** by combining the first and second received data sets. At step **216**, the method **200** includes, receiving the query from the first agent. Moreover, the coordinator **104** is configured to receive the first query from the first agent **102A**. At step **218**, the method **200** includes, determining a prompt for query based on the global scene model. Moreover, the third controller **128** of the coordinator **104** is configured to determine the prompt for the query received from the first agent **102A** based on the model of the global scene **132**. At step **220**, the method **200** includes, transmitting a prompt to the first agent. Moreover, the coordinator **104** is configured to transmit the prompt to the first agent **102A** through the communication network **106**. At step **222**, the method **200** includes, receiving a prompt. Moreover, the first controller **108** of the first agent **102A** is configured to receive the prompt from the coordinator. At step **224**, the method **200** includes reevaluating sensor data in view of the received prompt. Moreover, the first agent **102A** is configured to reevaluate the first low-resolution sensor data set **116** after receiving the prompt. At step **226**, the method **200** includes, performing the desired action. Moreover, the first agent **102A**, after reevaluating the sensor data according to the prompt, is configured to perform the desired action.

Advantageously, the method **200** is used to generate the comprehensive global scene model **132** of the dynamic environment based on the diverse local views of the autonomous agents. Furthermore, the generated global scene model **132** is utilized to enhance the decision-making abilities of the one or more agents **102**. Each agent is equipped with sensors and controllers that are configured to capture the image data, which can be further analysed to perform an action. Moreover, the method **200** is used to combine the data sets (i.e., the segmented data set and the low-resolution sensor data set) from the one or more agents **102** of the agent network **100** in order to generate the global scene model **132**. The generation of the global scene model **132** enables the one or more agents **102** of the agent network **100** to take informed decisions based on a collective understanding of the agent environment, leading to an efficient, effective, and coordinated actions. Moreover, the one or more agents **102** of the agent network **100** are free to join or leave the agent network **100** based on the requirement due to which the agent network **100** offers high flexibility. The one or more agents **102** in the method **200** contribute to a high bandwidth optimization by transmitting the data to the other one or more agents **102** and the coordinator **104**, when required in order to minimize the amount of data exchanged within the agent network **100** and enhances the optimized resource utilization. Furthermore, the one or more agents **102** enhances the privacy of the agent network **100** by selectively sending the segmented data and the low-resolution sensor data set to the coordinator **104** of the one or more agents **102** instead of directly sharing the raw data with the coordinator **104** of the one or more agents **102**. Moreover, the determination of the prompt for the query based on the global scene model **132** allows the one or more agents **102** of the method **200** to seek additional information that is further utilized to perform the desired action efficiently and reliably. As a result, the agent network **100** is configured to provide a scalable, resource-efficient, and reliable solution for accurate and efficient decision-making by the one or more agents **102** of Advantageously, the method **200** is configured to generate a comprehensive global scene model **132** of the dynamic environment based on the diverse local views of the autonomous agents. Furthermore, utilize the generated global scene model **132** to enhance the decision-making abilities of the one or more agents **102**. Each agent is equipped with sensors and controllers that are configured to capture the image data, which can be further analysed to perform an action. Moreover, the method **200** includes the coordinator **104** that is configured to combine the data sets (i.e., the segmented data set and the low-resolution sensor data set) from the one or more agents **102** of the agent network **100** in order to generate the global scene model **132**. The generation of the global scene model **132** enables the one or more agents **102** in the method **200** to take informed decisions based on a collective

understanding of the agent environment, leading to an efficient, effective, and coordinated actions. Moreover, the one or more agents **102** are free to join or leave the agent network **100** based on the requirements due to which the agent network **100** offers high flexibility. The one or more agents **102** in the method **200** contribute to a high bandwidth optimization by transmitting the data to the other one or more agents **102** and the coordinator **104**, when required in order to minimize the amount of data exchanged within the agent network **100** and enhances the optimized resource utilization. Furthermore, the one or more agents **102** enhances the privacy of the agent network **100** by selectively sending the segmented data and the low-resolution sensor data set to the coordinator **104** of the one or more agents **102** instead of directly sharing the raw data with the coordinator **104** of the one or more agents **102**. Moreover, the determination of the prompt for the query based on the global scene model **132** allows the one or more agents **102** of the agent network **100** to seek additional information that is further utilized to perform the desired action efficiently and reliably. As a result, the method **200** is used to provide a scalable, resource-efficient, and reliable solution for accurate and efficient decision-making by the one or more agents **102** of the method **200**.

The steps **202** to **226** are only illustrative, and other alternatives can also be provided where one or more steps are added, one or more steps are removed, or one or more steps are provided in a different sequence without departing from the scope of the claims herein.

There is further provided a computer program product comprising program instructions for performing the method **200** when executed by one or more processors in the agent network **100**. The computer program product is implemented as an algorithm, embedded in software stored in a non-transitory computer-readable storage medium. The non-transitory computer-readable storage means may include but are not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. Examples of implementation of computer-readable storage medium, but are not limited to, Electrically Erasable Programmable Read-Only Memory (EEPROM), Random Access Memory (RAM), Read Only Memory (ROM), Hard Disk Drive (HDD), Flash memory, a Secure Digital (SD) card, Solid-State Drive (SSD), a computer-readable storage medium, and/or CPU cache memory.

FIG. 3 is a block diagram that depicts an agent in an agent network comprising one or more agents, in accordance with an embodiment of the present disclosure. FIG. 3 is described in conjunction with elements from FIG. 1. With reference to FIG. 3, there is shown a diagram **300** that includes an agent (i.e., the first agent **102A of FIG.1)**, which includes the first controller **108**, the first sensor **110**, and the first memory **112**. Furthermore, the first memory includes the first segmented data set **114** and the first low-resolution sensor data set **116**.

There is provided an agent (i.e., the first agent **102A** of FIG. 1) in the agent network **100** comprising the one or more agents **102** and the coordinator **104**. However, an agent in the agent network **100** may correspond to the first agent **102A**, the second agent **102B**, up to nth agents without affecting the scope of the present disclosure. Each agent, such as the first agent **102A** includes a sensor (i.e., the first sensor **110**) and a controller (i.e., the first controller **108**). The first controller **108** is configured to receive sensor input from the sensor (i.e., the first sensor **110**) of a partial scene and identify features relevant for solving an action. The received sensor input from each of the one or more agents **102** is used to provide a comprehensive and detailed understanding of the dynamic environment in real-time that can be further utilized for recognizing objects, patterns, or any other relevant elements in the dynamic environment. The first controller **108** is configured to generate a segmented data set (i.e., the first segmented data set **114**) and a low-

resolution sensor data set (i.e., the first low-resolution sensor data set **116**). The generation of the segmented data set (e.g., the first segmented data set **114** and the second segmented data set **124**) and low-resolution sensor data set (i.e., the first low-resolution sensor data set **116** and the second low-resolution sensor data set **126**) are used to get an insight on a specific aspect of the partial scene of the dynamic environment. Furthermore, the first controller **108** is configured to send the first segmented data set **114** and the first low-resolution sensor data set **116** to the coordinator **104**. As a result, the generation of the first segmented data set **114** and the first low-resolution sensor data set **116** are transmitted to the coordinator **104** in order to provide a holistic representation of the dynamic environment.

Furthermore, the first controller **108** is configured to send a query to the coordinator **104**, receive a prompt based on a global scene model **132** for the query, reevaluate the sensor data in the view of the received prompt, and perform a desired action. Firstly, the first agent **102A** is configured to collect the local data and further evaluate if the local data has enough information to perform a desired action or not. Moreover, if the local information is insufficient, then, in that case, the first agent **102A** is configured to select a query (i.e., the query from a pre-defined query-list) that describes the required missing information. In an example, the received query forms the first agent **102A** includes "Can I safely turn right?", "Are there some obstacles ahead?", and the like. Thereafter, the first agent **102A** is configured to send the query that includes a tuple having an agent-id, a timestamp, a location-agent, and a query-id to the coordinator **104**. In an implementation, the query includes information about objects, attributes and relations for objects, topographical information, and additional background information. Moreover, the query is accompanied by locally processed data (e.g., Im-Seg, Im-LowRes, and the like). After that, the coordinator **104**, upon receiving the query and the local data from the first agent **102A**, is configured to extract all the information from the global scene model **132** (or a global database) that is related to the query and the received local data, such as information about objects, their attributes and relations, some topographical information of streets and buildings, and additional background information, for example, bad weather conditions by using data-retrieval techniques. Furthermore, the prompt includes an id-agent, a timestamp, a query-id, and a prompt. The first agent **102A** upon receiving the prompt is configured to re-evaluate its own objective by considering newly received information. As a result, the first controller **108** is configured to ensure an efficient and accurate decision-making process of the first agent **102A** thereby, enhancing the adaptability and responsiveness of the first agent **102A** in a dynamic environment contributing to the overall efficacy of the agent network **100**.

FIG. 4 is the diagram that depicts a flow chart for the method of an agent in an agent network comprising one or more agents, in accordance with an embodiment of the present disclosure. With reference to FIG. 4, there is shown a flowchart of method **400** of the first agent in the agent network. The method **400** includes steps **402** to **416**.

At step **402**, the method **400** includes, receiving sensor input from the sensor of a partial scene. Moreover, the first agent **102A** is configured to receive the sensor input of the partial scene. At step **404**, the method **400** includes, identifying features relevant for solving action. Moreover, the first agent **102A** is configured to identify the relevant features from the sensor input of the partial scene. At step **406**, the method **400** includes, generating segmented data sets and low-resolution data. Moreover, the first controller **108** of the first agent **102A** is configured to generate the first segmented data set **114** and the first low-resolution sensor data set **116**. At step **408**, the method **400** includes, sending a segmented data set and low-resolution sensor data set to the coordinator. Moreover, the first agent **102A** is configured to send the first segmented data set **114** and the first low-resolution sensor data set **116** to the coordinator **104** through the communication network **106**. At step **410**, the method **400** includes, sending a query to the coordinator

104. Moreover, the first agent **102A** is configured to send the query to the coordinator **104**. At step **412**, the method **400** includes, receiving prompts based on the global scene for the query. Moreover, the first agent **102A** is configured to receive the prompt based on the global scene model **132** of the query. At step **414**, the method **400** includes reevaluating sensor data in view of the received prompt. Moreover, the first agent **102A** is configured to reevaluate the sensor data according to the received prompt. At step **416**, the method **400** includes, performing the desired action. Moreover, the first agent **102A** is configured to perform the desired action. Advantageously, the method **400** is used to ensure an efficient and accurate decision-making process of the first agent **102A** thereby, enhancing the adaptability and responsiveness of the first agent **102A** in a dynamic environment contributing to the overall efficacy of the agent network **100**.

The steps **402** to **416** are only illustrative, and other alternatives can also be provided where one or more steps are added, one or more steps are removed, or one or more steps are provided in a different sequence without departing from the scope of the claims herein.

There is further provided a computer program product comprising program instructions for performing the method **400** when executed by one or more processors in the agent network **100**. The computer program product is implemented as an algorithm, embedded in software stored in a non-transitory computer-readable storage medium. The non-transitory computer-readable storage means may include but are not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. Examples of implementation of computer-readable storage medium, but are not limited to, Electrically Erasable Programmable Read-Only Memory (EEPROM), Random Access Memory (RAM), Read Only Memory (ROM), Hard Disk Drive (HDD), Flash memory, a Secure Digital (SD) card, Solid-State Drive (SSD), a computer-readable storage medium, and/or CPU cache memory.

FIG. 5 is a block diagram that depicts a coordinator in an agent network comprising one or more agents, in accordance with an embodiment of the present disclosure. FIG. 5 is described in conjunction with elements from FIG. 1 and FIG. 3. With reference to FIG. 5, there is shown a diagram **500** depicting the coordinator **104** (of FIG.1), that includes the global scene model **132** (of FIG.1), the third controller **128** (of FIG. 1) and the network interface **134** (of FIG.1).

There is provided the coordinator **104** in the agent network **100** including the one or more agents **102**. Moreover, each agent (e.g., the first agent **102A** and the second agent **102B**) comprises a sensor (e.g., the first sensor **110** and the second sensor **120)**. The coordinator **104** includes a controller (i.e., the third controller **128)** that is configured to receive the first segmented data set **114** and the first low-resolution sensor data set **116** from the first agent **102A**. Thereafter, the coordinator **104** is configured to receive the second segmented data set **124** and the second low-resolution sensor data set **126** from the second agent **102B**. The coordinator **104** is further configured to generate the global scene model **132** based on a combination of the received data sets from the agents. Furthermore, the coordinator **104** is configured to receive a query from the first agent **102A**, determine a prompt for the query based on the global scene model **132**, and transmit the prompt to the first agent (**102A**). Furthermore, after receiving the first segmented data set **114**, the first low-resolution sensor data set **116**, the second segmented data set **124**, and the second low-resolution sensor data set **126.** The third controller **128** is configured to generate a global scene model **132** based on a combination of the received data sets from the agents (i.e., from the first agent **102A** and the second agent **102B**). For example, the third controller **128** of the coordinator **104** is configured to receive the processed image data from the

agent and combine the processed data with information received from other agents. The arrangement of the combined image data is organized using received meta-information such as agent-id, timestamp, and location-agent. Furthermore, the combined data are processed consecutively by two or more algorithms, such as a cross-view attention neural network, a diffusion attention neural network, and like without limiting the scope of the present disclosure. Advantageously, the generation of the comprehensive global scene model **132** serves as a knowledge database, and the third controller **128** handles queries, determining prompts for informed decision-making, effectively increasing the scalability and adaptability of the one or more agents in the agent network.

FIG. 6 is the diagram that depicts a flow chart for the method of an agent in an agent network, in accordance with an embodiment of the present disclosure. With reference to FIG. 6, there is shown a flowchart of method **600** of the agent (i.e., the first agent **102A**) in the agent network **100**.

At step **602**, the method **600** includes the coordinator **104 for** receiving the first segmented data set **114** and the first low-resolution sensor data set **116** from the first agent **102A**. At step **604**, the method **600** includes receiving the second segmented data set **124** and the second low-resolution sensor data set **1**26 from the second agent **102B**. Thereafter, at step **606**, the method **600** includes generating a model of a global scene **132** based on a combination of the received data sets from the agents. After that, at step **608**, the method **600** includes receiving a query from the first agent **102A**, and at step **610**, the method **600** includes determining a prompt for the query based on the global scene model. Finally, at step **612**, the method **600** includes transmitting the prompt to the first agent. Advantageously, the generation of the comprehensive global scene model **132** serves as a knowledge database, and the third controller **128** handles queries, determining prompts for informed decision-making, effectively increasing the scalability and adaptability of the one or more agents in the agent network.

The steps **602** to **612** are only illustrative, and other alternatives can also be provided where one or more steps are added, one or more steps are removed, or one or more steps are provided in a different sequence without departing from the scope of the claims herein.

There is further provided a computer program product comprising program instructions for performing the method **600** when executed by one or more processors in the agent network **100**. The computer program product is implemented as an algorithm, embedded in software stored in a non-transitory computer-readable storage medium. The non-transitory computer-readable storage means may include but are not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. Examples of implementation of computer-readable storage medium, but are not limited to, Electrically Erasable Programmable Read-Only Memory (EEPROM), Random Access Memory (RAM), Read Only Memory (ROM), Hard Disk Drive (HDD), Flash memory, a Secure Digital (SD) card, Solid-State Drive (SSD), a computer-readable storage medium, and/or CPU cache memory.

FIG. 7A is a diagram depicting a recreation of the global scene model, in accordance with an embodiment of the present disclosure. FIG. 7A is described in conjunction with elements from FIGs. 1, 3, and 5. With reference to FIG. 7A, there is shown a diagram **700A** depicting an agent (i.e., the first agent **102A** of FIG.1) and the coordinator **104** (of FIG.1).

In an implementation scenario, each agent (e.g., the first agent **102A**) is equipped with the first sensor **110**, such as a camera sensor, which provides a partial understanding of the dynamic environment. In addition, each agent, such as the first agent **102A** performs an action **702** to perform regeneration of the global scene model **132** and query-prompt knowledge sharing. The first agent **102A** is configured to capture a snapshot of a local view (i.e., an image **704** from a camera). The captured data is then processed by a task-agnostic compression algorithm that removes all the redundancy from the sensing data and a take-aware processing algorithm that emphasizes all information related to the agent's task while suppressing unrelated information. Moreover, the output of the implementation of the task-agnostic compression algorithm is a low-resolution version of the input view Im-LowRes and the output of the implementation of the task-aware processing algorithm is a segmented image (i.e., Im-Seg) which illustrates the contours of detected relevant detected objects (e.g., cars, pedestrians, street lines, and the like.). The first agent **102A** is configured to use vision neural network **706** to process the sensing data and send the processed input data to the coordinator **104**, such as at operation **714**. Furthermore, the coordinator **104** is configured to receive the processed sensing data that includes agent-id, timestamp, location-agent, Im-Seg, and Im-LowRes. Furthermore, the first agent **102A** is configured to generate a list of queries **708** and send an action-dependent query **726**, such as at operation **720** to the coordinator **104**. Furthermore, the coordinator **104** is configured to generate a prompt **722** based on the action-dependent query through a prompt generation operation **718**. Moreover, the coordinator **104** is further configured to send the action-dependent query **726** along with an action-dependent prompt to the first agent **102A**, such as at operation **724**. In addition, the first agent **102A** executes an artificial neural network **710** to perform the desired action. At operation **716**, the coordinator **104** receives the processed data from the first agent **102A** and concatenates the processed data with the data received from other agents of the one or more agents **102**. Moreover, the ordering of the concatenated data is maintained by the received meta-information (i.e., agent-id, timestamp, location-agent). Furthermore, the concatenated data are processed consecutively by using transformers, such as by executing a cross-view attention neural network and a diffusion attention neural network. The cross-view attention model is trained to associate (at a semantic level) pieces of data provided by the one or more agents **102**. Such association may include identifying objects seen by the one or more agents **102** at different angles, discovering relations between two distant objects observed by different agents, and positioning the views within the global 3D coordinate system. In addition, the diffusion attention neural network is trained to regenerate the global scene model **132** from the output of the cross-view attention neural network and from the received data from other agents, such as at operation **712**. The diffusion attention neural network is used to produce new synthetic information that may fill the gaps to allow the regeneration of the global scene model **132** suitable for machine-type processing and communication.

FIG. 7B is a diagram depicting a local view processed by an agent, in accordance with an embodiment of the present disclosure. FIG. 7B is described in conjunction with elements from FIGs. 1, 3, 5, and 7A. With reference to FIG. 7B, there is shown a diagram **700B** depicting the local view processed by the agent, such as the first agent **102A** or the second agent **102B** from the one or more agents **102**. In an implementation scenario, there is shown an original view **728A**, a low-resolution version **728B** (i.e., Im-LowRes) of the original view, and a segmented version **728C** (i.e., Im-Seg) of the original view. As a result, the one or more agents **102** are configured to take informed decisions by accurately and reliably viewing the dynamic environment in real-time without any latency.

FIG. 8 is a diagram depicting an input processing of the agent, in accordance with an embodiment of the present disclosure. With reference to FIG. 8, there is shown a diagram **800** depicting an input image **802**, based on which the agent performs a desired action, such as at operation **804**.

At operation **806**, when the agent is not able to perform the desired action, the first controller **108** of the first agent **102A** is configured to perform a task-aware segmentation **808** and a task-agnostic image compression **810**. Moreover, at the task-aware segmentation **808**, the input image **802** undergoes a full image segmentation **814** to generate a segmented image **816** and, the segmented image **816** undergoes a goal-aware filtering **818** to generate a task-filtered segmented image **822**. Thereafter, the first agent **102A** from the one or more agents **102** is configured to perform full image segmentation using a trained neural network, such as Mask R-CNN. Concurrently, the input image **802** is compressed using some task-agnostic image compression technique. Furthermore, at the task-agnostic image compression **810**, the input image **802** undergoes lossy image compression **820** to generate a low-resolution image **824**. Moreover, the task-filtered segmented image **822** and the low-resolution image **824** undergo a positional embedding **812** that allows the first agent **102A** to take the desired necessary actions in real-time scenarios.

FIG. 9 is a diagram that illustrates global scene model regeneration using cross-view attention and diffusion, in accordance with an embodiment of the present disclosure. With reference to FIG. 9, there is shown allow-resolution and a segmented image data with positional embeddings **902** received from the one or more agents **102** that undergoes artificial neural network-based data fusion **904** and combines with a global map embedding **908** using queries **910**. The low-resolution and segmented image data with the positional embeddings **902** undergoes a cross-view artificial neural network **912**. The cross-view artificial neural network **912** receives three types of inputs such as queries, keys, and values. The keys and values are the combined tuples (such as Im-Seg, Im-LowRes, embedding) from multiple agents. The queries are a global map embedding which could be viewed as an empty model structure of the global scene to be filled with semantic information.

At operation **906**, the low-resolution and segmented image data with the positional embeddings **902** then undergoes diffusion-based scene regeneration a diffusion artificial neural network **914** to regenerate the global scene model **132**. The low-resolution and segmented image data with the positional embeddings **902** and the cross-view artificial neural network **912** are associated with each other and all relevant semantic information is filled in the global map embedding **908**. The resulting partially filled model is then fed to the diffusion artificial neural network **914** for training. The diffusion artificial neural network **914** is configured to fill the information gaps. The output of the diffusion artificial neural network **914** is a complete and the global scene model **132**, viewed as a database **916**. As a result, the global scene model **132** regeneration using cross-view and diffusion network model is performed.

FIG. 10 is a diagram that illustrates a prompt generation by a coordinator, in accordance with an embodiment of the present disclosure. FIG. 10 is described in conjunction with elements from FIGs. 1, 3, 5, 7A, 8, and 9. With reference to FIG.10, there is shown a diagram **1000** depicting the coordinator **104** and the global scene model **132**. Moreover, the global scene model **132** includes the database for the model that includes attributes for example, id, class, attr., and the like. As shown, the coordinator **104** is configured to extract from the global scene model **132** facts and features relevant to the first agent **102A** upon receiving the first agent's query and its processed local views. The facts and features include information about objects, their attributes and relations, some topographical information such as streets and buildings, and additional background information such as bad weather conditions. The extraction of relevant information is performed using conventional data-retrieval techniques. As a result, based on the relevant information, the first agent **102A** is configured to take the desired action.

FIG. 11 is a diagram that illustrates a scene-aware semantic analysis at the end of an agent, in accordance with an embodiment of the present disclosure. With reference to FIG.11, there is shown an agent **1100** (e.g., the first agent **102A** or the second agent **102B**), after receiving a prompt **1102A**, the agent undergoes a prompt-aware image segmentation **1104**, where the prompt **1102A** undergoes a vectorization **1106** to generate a scene map features vector **1108** and an image **1102B** undergoes a regional proposal network **1110** to generate object feature vector **1112**. Moreover, the scene map features vector **1108** and the object feature vector **1112** are concatenated to generate a cross attention **1114** to generate refined feature vector **1116**. Furthermore, the refined feature vector helps the agent **1100** to perform an action **1118** accurately and efficiently.

Modifications to embodiments of the present disclosure described in the foregoing are possible without departing from the scope of the present disclosure as defined by the accompanying claims. Expressions such as “including”, “comprising”, “incorporating”, “have”, “is” used to describe, and claim the present disclosure are intended to be construed in a non-exclusive manner, namely allowing for items, components or elements not explicitly described also to be present. Reference to the singular is also to be construed to relate to the plural. The word “exemplary” is used herein to mean “serving as an example, instance, or illustration”. Any embodiment described as “exemplary” is not necessarily to be construed as preferred or advantageous over other embodiments or to exclude the incorporation of features from other embodiments. The word “optionally” is used herein to mean “is provided in some embodiments and not provided in other embodiments”. It is appreciated that certain features of the present disclosure, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable combination or as suitable in any other described embodiment of the disclosure.

CLAIMS

1. An agent network (100) comprising one or more agents (102) and a coordinator (104), wherein each agent comprises a sensor and a controller configured to

receive sensor input from the sensor of a partial scene,

identify features relevant for solving an action,

generate a segmented data set and a low-resolution sensor data set and

send the segmented data set and the low-resolution sensor data set to the coordinator (104), and wherein the coordinator (104) comprises a controller configured to

receive a first segmented data set (114) and a first low-resolution sensor data set (116) from a first agent (102A),

receive a second segmented data set (124) and a second low-resolution sensor data set (126) from a second agent (102B),

generate a global scene model (132) based on a combination of the received data sets from the agents,

receive a query from the first agent,

determine a prompt for the query based on the model of a global scene (132), and

transmit the prompt to the first agent (102A), whereby

the controller of the first agent (102A) is further configured to

receive the prompt,

reevaluate the sensor data in the view of received prompt, and

perform a desired action.

2. The agent network (100) according to claim 1, wherein the controller of the first agent (102A) is further configured to reevaluate the sensor data in the view of received prompt by

assessing the transformed prompt combined with the identified features jointly using a neural network providing an enriched set of features, and

selecting the desired action based on the enriched set of features.

3. The agent network (100) according to claim 1 or 2, wherein the controller of the first agent is further configured to

determine that the sensor input is insufficient to solve the action, and in response thereto send the query to the coordinator (104).

4. The agent network (100) according to any preceding claim, wherein the controller of the first agent (102A) is further configured to

determine that the sensor input is insufficient to solve the action, and in response thereto perform full segmentation using a trained neural network,

filter out segmented objects unrelated to a local objective of the task, and

transmit positional information about the agent's global position and orientation, and a timestamp indicating a moment of recording to the coordinator (104).

5. The agent network (100) according to any preceding claim, wherein the controller of the coordinator (104) is further configured to generate the model of the global scene (132) based on a combination of the received data sets from the agents by running cross-view attention and diffusion.

6. The agent network (100) according to any preceding claim, wherein the controller of the coordinator (104) is further configured to

determine the prompt for the query based on the model of a global scene (132), by extracting query-related data from the model of a global scene (132).

7. The agent network (100) according to any preceding claim, wherein the query-related data include one, some, or all of the information about objects, attributes and relations for objects, topographical information, and additional background information, and wherein the controller of the coordinator (104) is further configured to extract query-related data using data-retrieval techniques.

8. The agent network (100) according to any preceding claim, wherein the sensor is an image sensor and the sensor data is image data, or wherein the sensor is a RADAR sensor and the sensor data is RADAR image data.

9. The agent network (100) according to any preceding claim, wherein the controller of the first agent (102A) is further configured to identify features through image segmentation.

10. A method (200) for an agent network (100) comprising one or more agents (102) and a coordinator (104), wherein each agent comprises a sensor, wherein the method comprises the agent for

receiving sensor input from the sensor of a partial scene,

identifying features relevant for solving an action,

generating a segmented data set and a low-resolution sensor data set and

sending the segmented data set and the low-resolution sensor data set to the coordinator (104), and wherein the method comprises the coordinator (104),

receiving a first segmented data set (114) and a first low-resolution sensor data set (116) from a first agent (102A),

receiving a second segmented data set (124) and a second low-resolution sensor data set (126) from a second agent (102B),

generating a model of a global scene (132) based on a combination of the received data sets from the agents,

receiving a query from the first agent (102A),

determining a prompt for the query based on the model of a global scene (132), and

transmitting the prompt to the first agent (102A), whereby the method (200) further comprises:

the first agent (102A)

receiving the prompt,

reevaluating the sensor data in the view of received prompt, and

performing a desired action.

11. An agent in an agent network (100) comprising one or more agents (102A) and a coordinator (104), wherein each agent comprises a sensor and a controller configured to

receive sensor input from the sensor of a partial scene,

identify features relevant for solving an action,

generate a segmented data set and a low-resolution sensor data set and

send the segmented data set and the low-resolution sensor data set to the coordinator (104),

send a query to the coordinator (104),

receive a prompt based on a global scene for the query,

reevaluate the sensor data in the view of received prompt, and

perform a desired action.

12. A method (400) for an agent in an agent network (100) comprising one or more agents (102) and a coordinator (104), wherein each agent comprises a sensor, wherein the method (400) comprises the agent

receiving sensor input from the sensor of a partial scene,

identifying features relevant for solving an action,

generating a segmented data set and a low-resolution sensor data set and

sending the segmented data set and the low-resolution sensor data set to the coordinator (104),

sending a query to the coordinator (104),

receiving a prompt based on a global scene for the query,

reevaluating the sensor data in the view of received prompt, and

performing a desired action.

13. A coordinator (104) in an agent network (100) comprising one or more agents (102), wherein each agent comprises a sensor, wherein the coordinator (104) comprises a controller configured to

receive a first segmented data set (114) and a first low-resolution sensor data set (116) from a first agent (102A),

receive a second segmented data set (124) and a second low-resolution sensor data set (126) from a second agent (102B),

generate a model of a global scene (132) based on a combination of the received data sets from the agents,

receive a query from the first agent (102A),

determine a prompt for the query based on the global scene model (132), and

transmit the prompt to the first agent (102A).

14. A method (600) for coordinator (104) in an agent network (100) comprising one or more agents (102), wherein each agent comprises a sensor, wherein the method (600) comprises the coordinator (104)

receiving a first segmented data set (114) and a first low-resolution sensor data set (116) from a first agent (102A),

receiving a second segmented data set (124) and a second low-resolution sensor data set (126) from a second agent (102B),

generating a model of a global scene (132) based on a combination of the received data sets from the agents,

receiving a query from the first agent (102A),

determining a prompt for the query based on the global scene model (132), and

transmitting the prompt to the first agent (102A).

15. A computer program product comprising program instructions for performing the method (200,400, 600) according to claims 10, 12, or 14, when executed by one or more processors in an agent network (100).