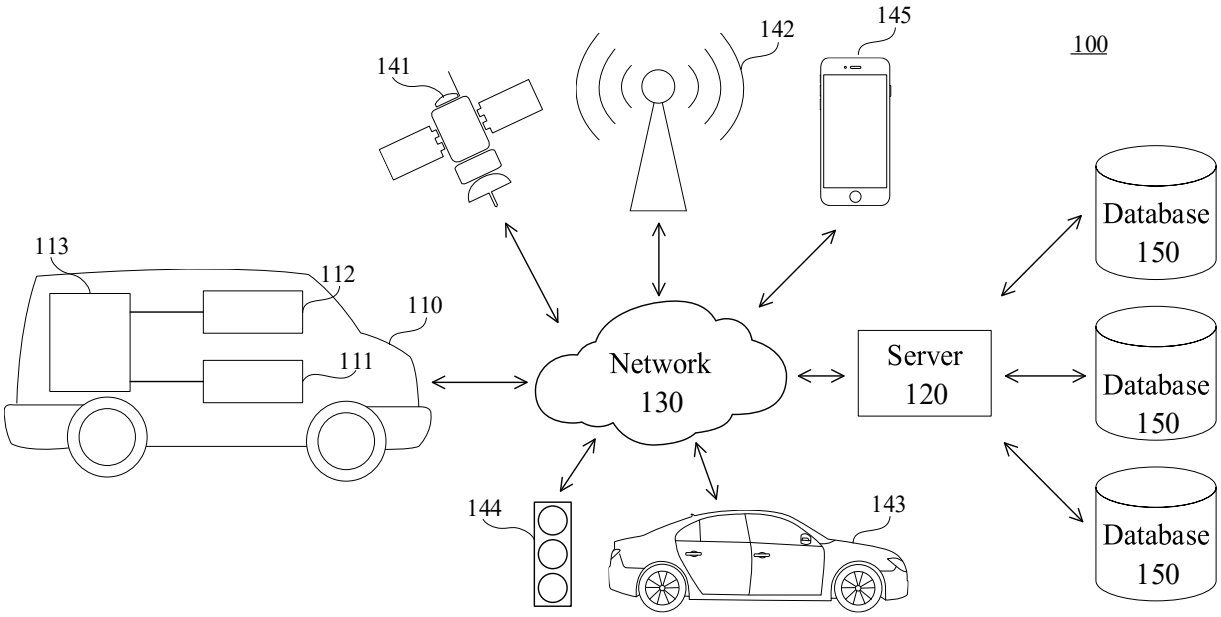

100
141
142
145
113
112
110
111
Network
130
Server
120
Database
150
Database
150
Database
150
144
143

FIG. 1

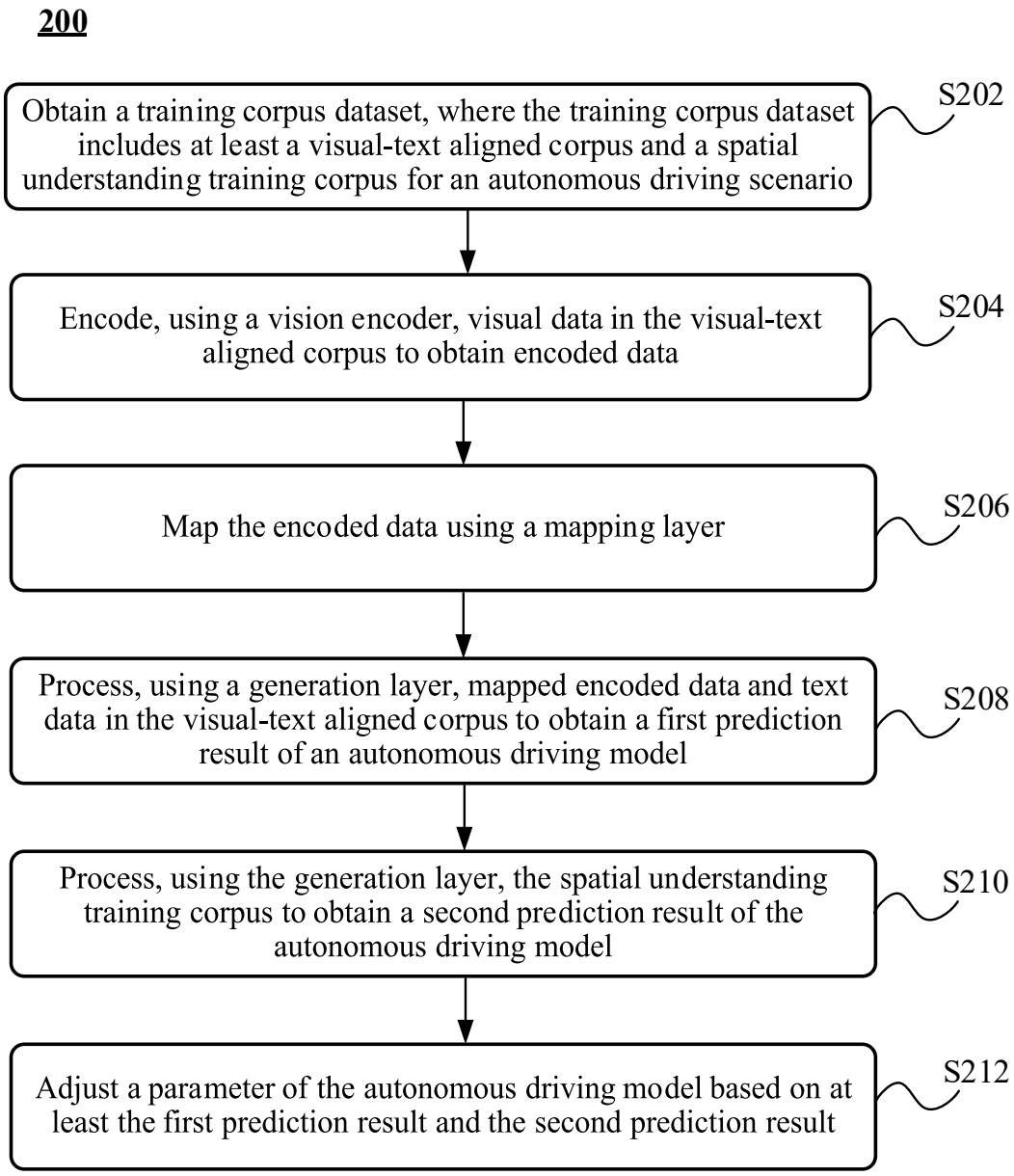

200
Obtain a training corpus dataset, where the training corpus dataset includes at least a visual-text aligned corpus and a spatial understanding training corpus for an autonomous driving scenario
S202
Encode, using a vision encoder, visual data in the visual-text aligned corpus to obtain encoded data
S204
Map the encoded data using a mapping layer
S206
Process, using a generation layer, mapped encoded data and text data in the visual-text aligned corpus to obtain a first prediction result of an autonomous driving model
S208
Process, using the generation layer, the spatial understanding training corpus to obtain a second prediction result of the autonomous driving model
S210
Adjust a parameter of the autonomous driving model based on at least the first prediction result and the second prediction result
S212

FIG. 2A

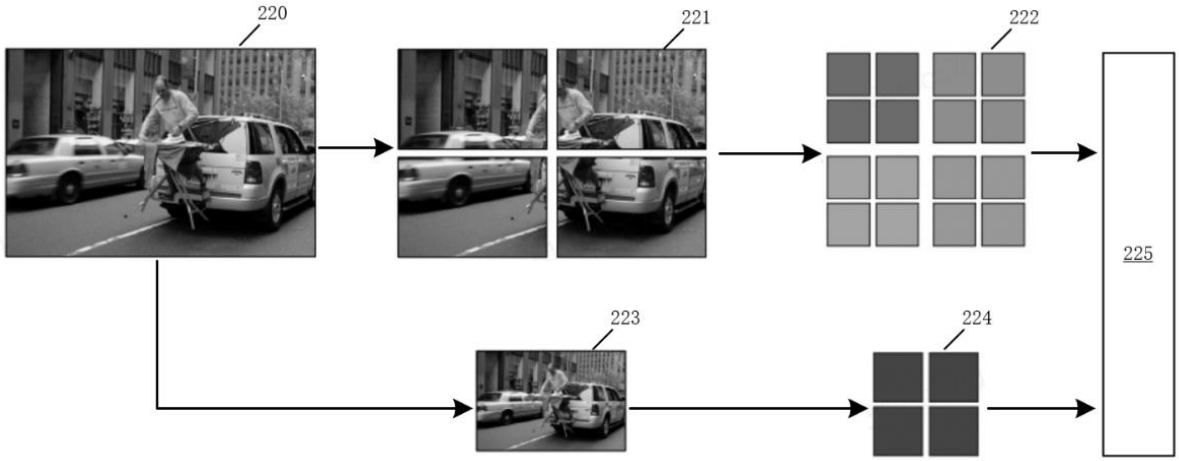


FIG. 2B

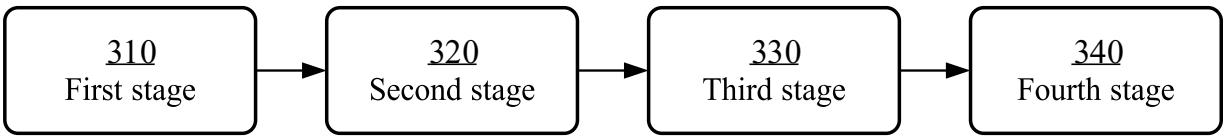


FIG. 3

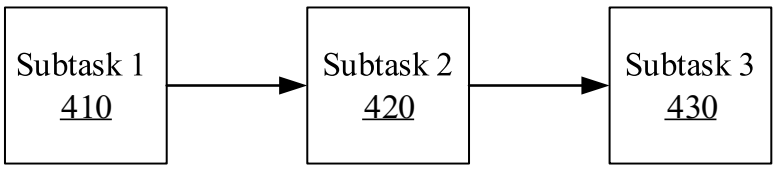


FIG. 4

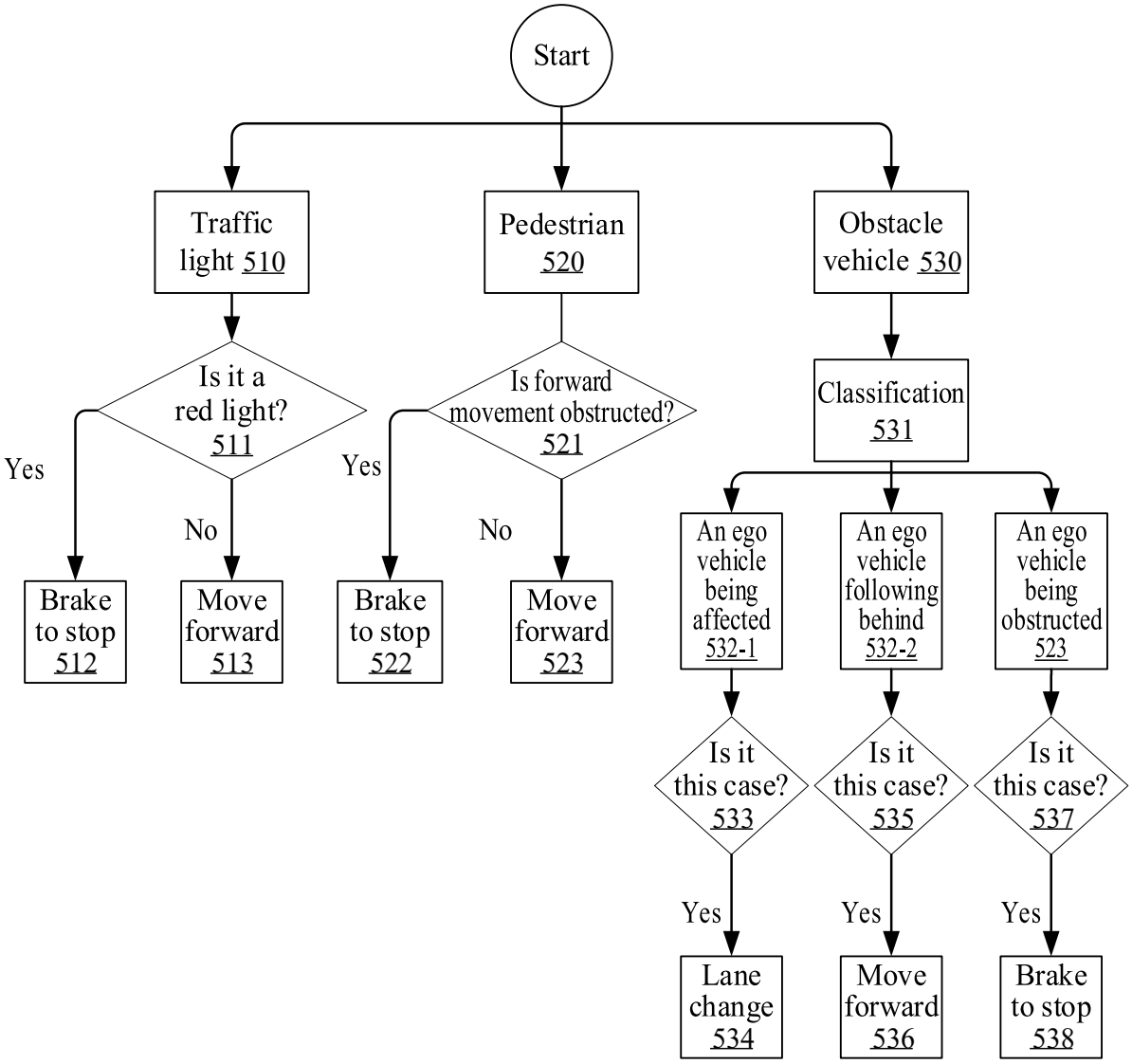

Start
Traffic light 510
Pedestrian 520
Obstacle vehicle 530
Is it a red light? 511
Yes
No
Brake to stop 512
Move forward 513
Is forward movement obstructed? 521
Yes
No
Brake to stop 522
Move forward 523
Classification 531
An ego vehicle being affected 532-1
An ego vehicle following behind 532-2
An ego vehicle being obstructed 523
Is it this case? 533
Is it this case? 535
Is it this case? 537
Yes
Yes
Yes
Lane change 534
Move forward 536
Brake to stop 538

FIG. 5


1
2
3
4
5
6

FIG. 6A

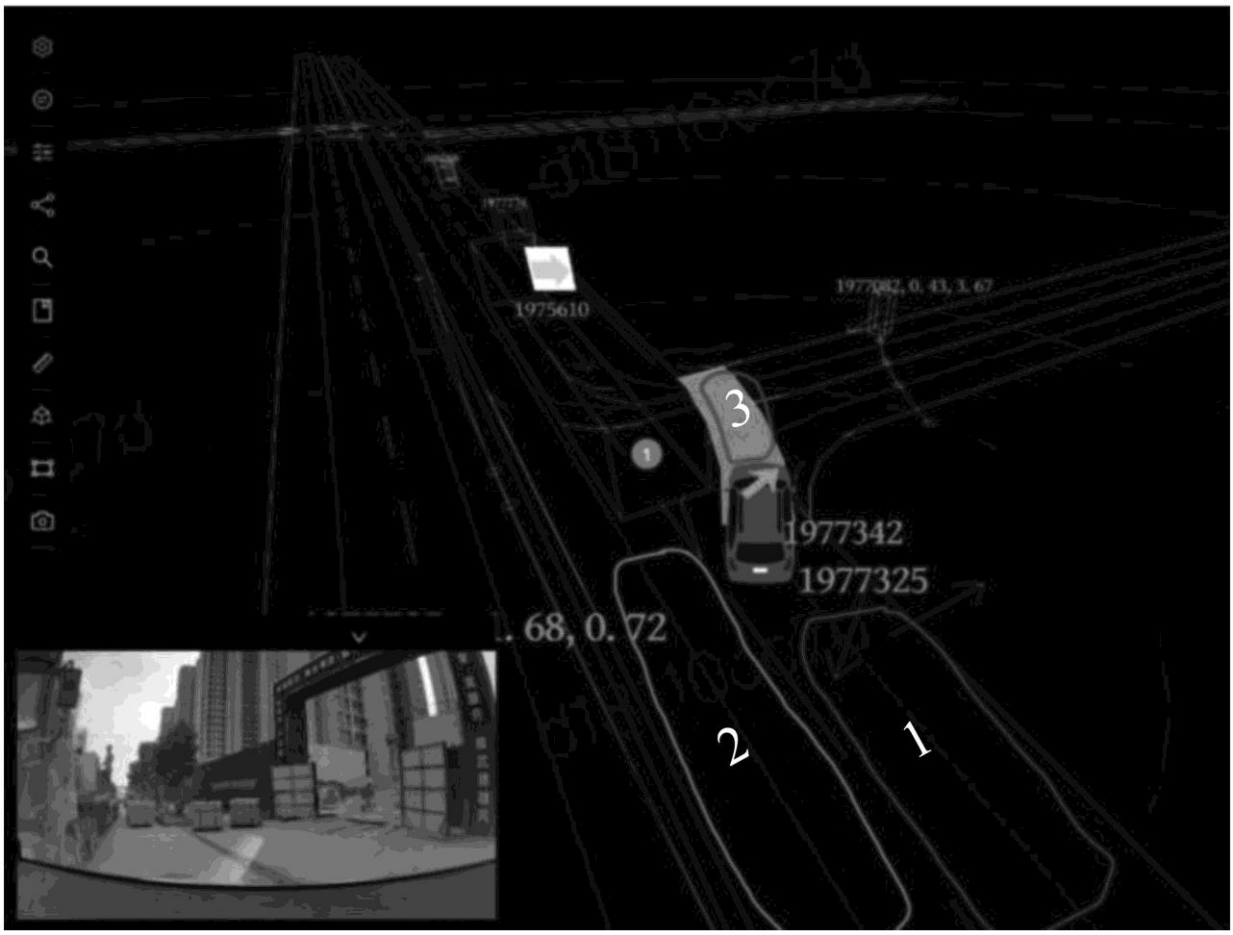

1975610
3
1977342
1977325
2
1

FIG. 6B

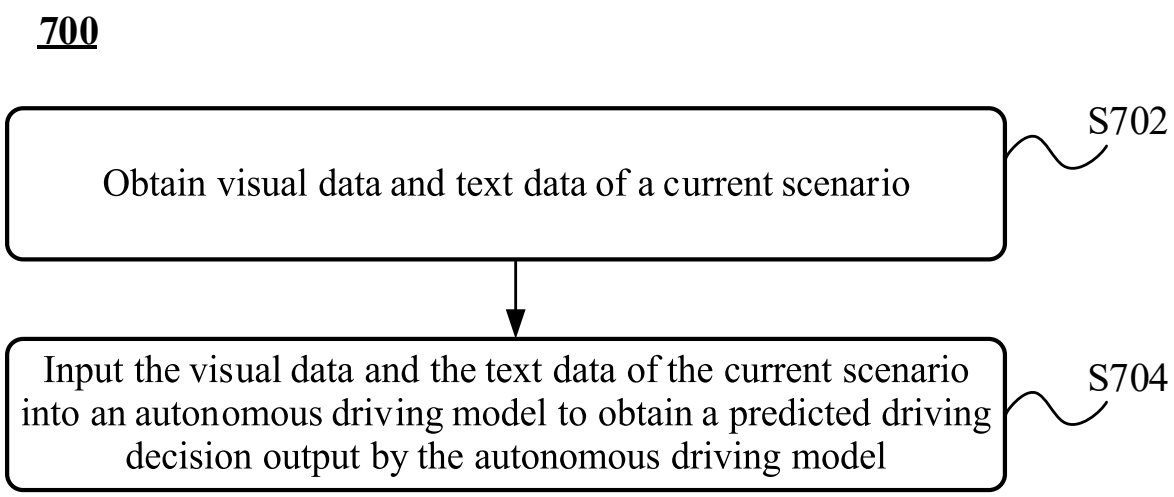

700
Obtain visual data and text data of a current scenario
S702
Input the visual data and the text data of the current scenario into an autonomous driving model to obtain a predicted driving decision output by the autonomous driving model
S704

FIG. 7A

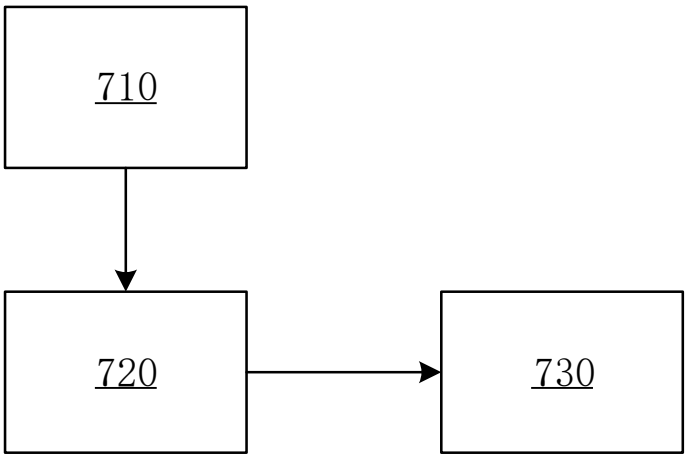


FIG. 7B

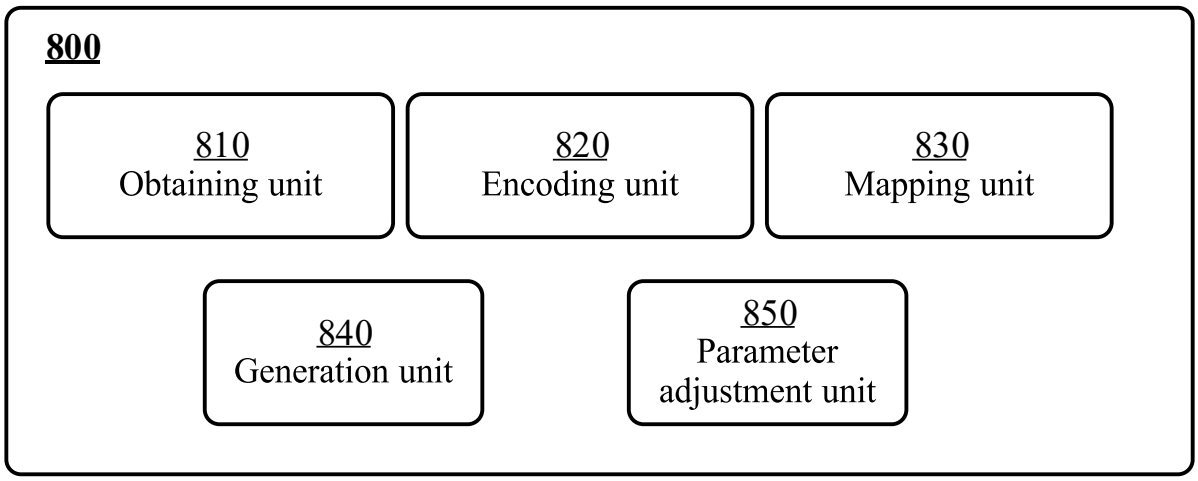


FIG. 8

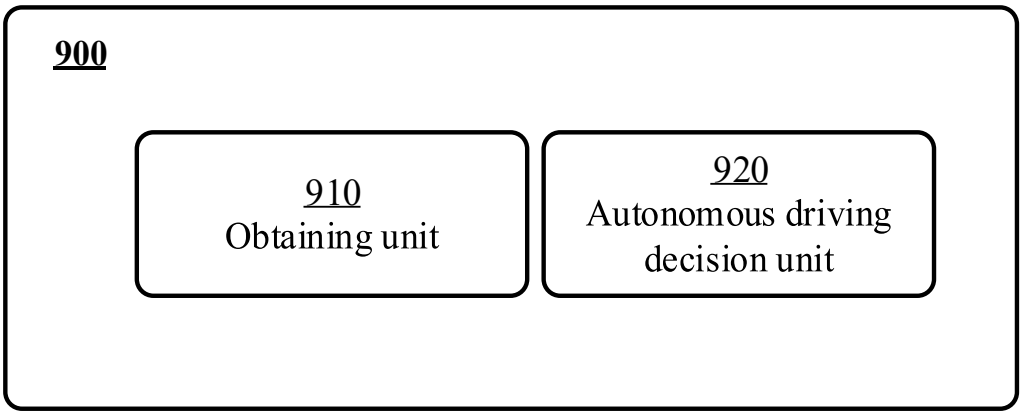


FIG. 9

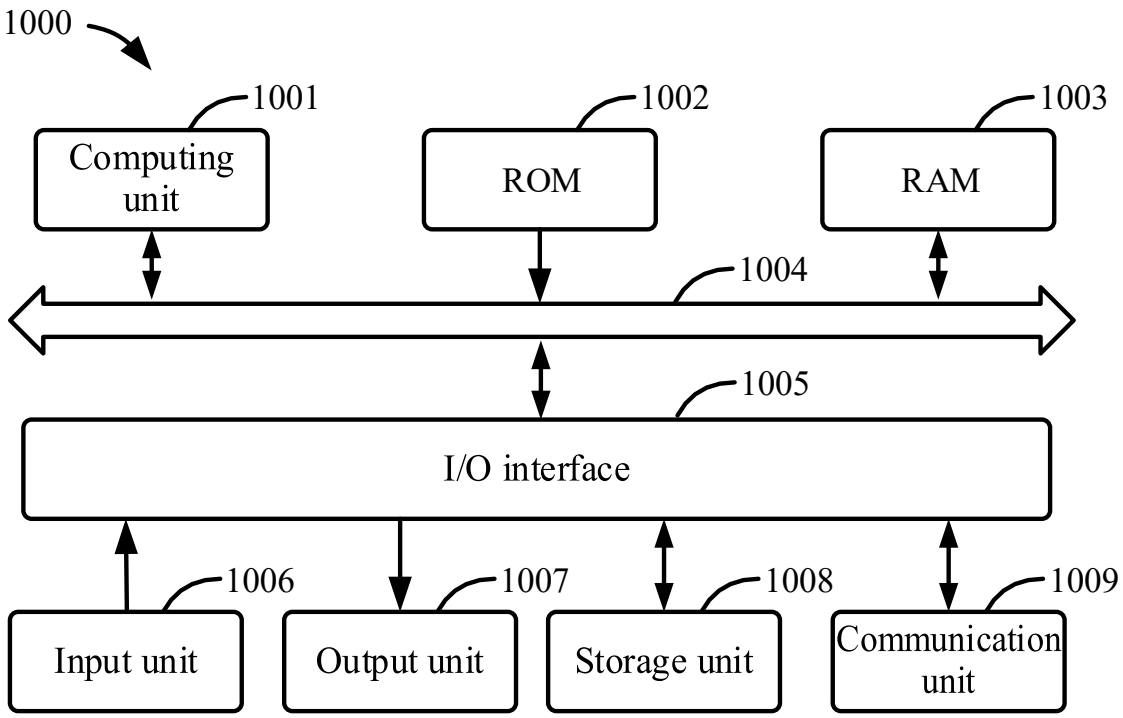

1000
1001
Computing unit
1002
ROM
1003
RAM
1004
1005
I/O interface
1006
Input unit
1007
Output unit
1008
Storage unit
1009
Communication unit

FIG. 10