(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 June 2022 (30.06.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/134996 A1

(51) International Patent Classification:
G06K 9/00 (2022.01)

(21) International Application Number:
PCT/CN2021/132554

(22) International Filing Date:
23 November 2021 (23.11.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
202011555482.9 25 December 2020 (25.12.2020) CN

(71) Applicant: **ZHEJIANG DAHUA TECHNOLOGY CO., LTD.** [CN/CN]; No. 1187 Binan Road, Binjiang District, Hangzhou, Zhejiang 310051 (CN).

(72) Inventor: **YANG, Xuefeng**; No. 1187 Binan Road, Binjiang District, Hangzhou, Zhejiang 310051 (CN).

(74) Agent: **CHINA WISPRO INTELLECTUAL PROPERTY LLP.**; Room A806, Zhongdi Building, China University of Geosciences Base, No.8 Yuexing 3rd Road, High-Tech Industrial Estate, Nanshan District, Shenzhen, Guangdong 518057 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: LANE LINE DETECTION METHOD BASED ON DEEP LEARNING, AND APPARATUS

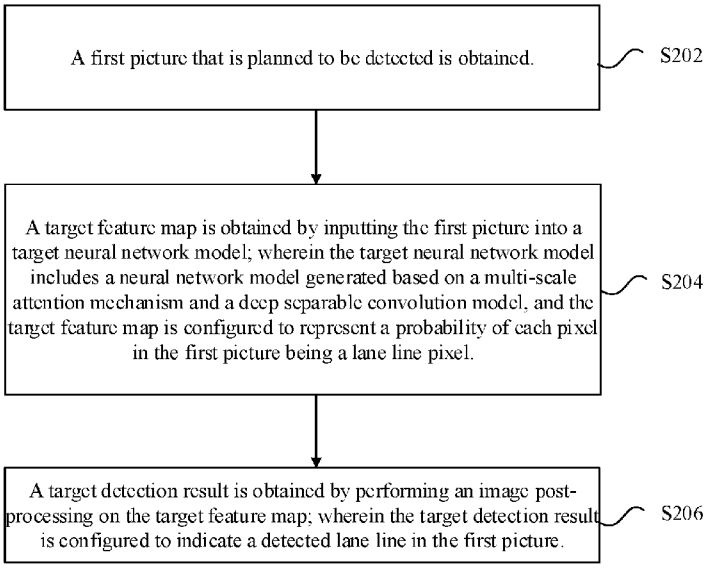


FIG. 2

(57) Abstract: Disclosed are a lane line detection method based on deep learning, an apparatus, a storage medium, and an electronic device. The method includes: obtaining a first picture that is planned to be detected; obtaining a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel; and obtaining a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

WO 2022/134996 A1

[Continued on next page]

SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

LANE LINE DETECTION METHOD BASED ON DEEP LEARNING, AND APPARATUS

CROSS REFERENCE

[0001] The present application claims foreign priority of China Patent Application No. 202011555482.9 filed on December 25, 2020, in the China National Intellectual Property Administration, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

[0002] The present disclosure relates to the field of communication technologies, and in particular to a lane line detection method based on deep learning, an apparatus, a storage medium, and an electronic device.

BACKGROUND

[0003] With the rapid development of technology and economy, there are more and more cars running on the road, which facilitates people's travel while bringing more and more traffic accidents and traffic congestion problems. Autonomous driving and an intelligent assisted driving system can help a driver process most of road information, provide precise guidance for the driver, and reduce the probability of traffic accidents. An intelligent transportation system can identify the number of vehicles on a lane, determine whether the lane is congested, plan a more reasonable travel route for the driver, and alleviate traffic congestion. In the automatic driving, intelligent driving assistance system and intelligent transportation system, a vision-based lane line detection is very critical and is the basis and core technology for realizing lane departure warning and lane congestion warning.

[0004] In the current related technologies, the method for lane line detection requires a large number of training samples and a complex neural network model, resulting in a technical problem that the efficiency of detecting lane lines is very low.

[0005] Aiming at the technical problem of low efficiency of detecting lane lines in the related technologies, no effective solution has been proposed yet.

SUMMARY OF THE DISCLOSURE

[0006] The present disclosure provides a lane line detection method based on deep learning, an apparatus, a storage medium, and an electronic device, to solve the technical problem of low detection accuracy of lane lines in the related art.

[0007] In a first aspect, the present disclosure provides a lane line detection method based on deep learning, comprising: obtaining a first picture that is planned to be detected; obtaining a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model comprises a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of

each pixel in the first picture being a lane line pixel; and obtaining a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[0008] In a second aspect, the present disclosure provides a lane line detection apparatus based on deep learning, comprising: an obtaining module, configured to obtain a first picture that is planned to be detected; a first processing module, configured to obtain a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model comprises a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel; and a second processing module, configured to obtain a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[0009] In a third aspect, the present disclosure provides a computer-readable storage medium, storing a computer program; wherein the computer program is configured to perform the method as described when executed.

[0010] In a fourth aspect, the present disclosure provides an electronic device, comprising a processor, a memory, and a computer program stored in the memory and executable on the processor; wherein the processor is configured to perform the method as described above when executing the computer program.

[0011] In the present disclosure, a first picture that is planned to be detected is obtained; a target feature map is obtained by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel; and a target detection result is obtained by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture. Therefore, the problem of low detection accuracy of lane lines in the related art may be solved, thereby increasing the detection efficiency of lane lines, increasing the detection accuracy of lane lines, and reducing the detection cost.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The drawings described here are to provide a further understanding of the present disclosure and constitute a part of the present disclosure. The exemplary embodiments and their descriptions of the present disclosure are to explain the present disclosure, and do not constitute an improper limitation of the present disclosure.

[0013] FIG. 1 is a block view of a hardware structure of a mobile terminal performing a lane line detection method based on deep learning according to an embodiment of the present disclosure.

[0014] FIG. 2 is a flowchart of a lane line detection method based on deep learning according to an embodiment of the present disclosure.

[0015] FIG. 3 is a schematic view of a lane line detection method based on deep learning according to an embodiment of the present disclosure.

[0016] FIG. 4 is a schematic view of a lane line detection method based on deep learning according to another embodiment of the present disclosure.

[0017] FIG. 4a is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0018] FIG. 4b is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0019] FIG. 4c is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0020] FIG. 5 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0021] FIG. 6 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0022] FIG. 7 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0023] FIG. 8 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0024] FIG. 9 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure.

[0025] FIG. 10 is a structural block view of a lane line detection apparatus based on deep learning according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0026] The embodiments of the present disclosure will be described in detail with reference to the drawings and in conjunction with the embodiments.

[0027] It should be noted that the terms “first” and “second” in the specification and claims of the present disclosure and the drawings are used to distinguish similar objects, and not necessarily to describe a specific sequence or order.

[0028] The method embodiments provided in the embodiments of the present application may be executed in a mobile terminal, a computer terminal or a similar computing device. Taking running on a mobile terminal as an example, FIG. 1 is a block view of a hardware structure of a mobile terminal performing a lane line detection method based on deep learning according to an embodiment of the present disclosure. As shown in FIG. 1, the mobile terminal may include one or more (only one is shown in FIG. 1) processor 102 (the processor 102 may include, but is not limited to, a processing device such as a microprocessor MCU or a programmable logic device FPGA) and a memory 104 for storing data. The mobile terminal may also include a transmission device 106 and an input/output device 108 for

communication functions. Those skilled in the art can understand that the structure shown in FIG. 1 is only for illustration and does not limit the structure of the mobile terminal. For example, the mobile terminal may also include more or fewer components than shown in FIG. 1, or have a different configuration from that shown in FIG.

[0029] The memory 104 may be configured to store computer programs, for example, software programs and modules of application software. Specifically, the memory may store computer programs corresponding to the lane line detection method based on deep learning herein the embodiments of the present disclosure. The processor 102 runs the computer programs stored in the memory 104 to perform various functional applications and data processing, that is, to achieve the above method. The memory 104 may include a high-speed random access memory, and may also include a non-volatile memory, such as one or more magnetic storage devices, flash memory, or other non-volatile solid-state memory. In some examples, the memory 104 may further include a memory remotely arranged relative to the processor 102, and the remote memory may be connected to the mobile terminal through a network. Examples of the network include, but are not limited to, the Internet, corporate intranet, local area network, mobile communication network, and any combination thereof.

[0030] The transmission device 106 is configured to receive or send data via a network. Specific examples of the network may include a wireless network provided by a communication provider of the mobile terminal. In some examples, the transmission device 106 includes a network interface adapter (NIC), which can be connected to other network devices through a base station to communicate with the Internet. In some examples, the transmission device 106 may be a radio frequency (RF) module configured to communicate with the Internet in a wireless manner.

[0031] In the embodiments, a lane line detection method based on deep learning running on a mobile terminal, a computer terminal or a similar computing device is provided. FIG. 2 is a flowchart of a lane line detection method based on deep learning according to an embodiment of the present disclosure. The method may include operations at blocks as followed.

[0032] At block S202: A first picture that is planned to be detected is obtained.

[0033] At block S204: A target feature map is obtained by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel.

[0034] At block S206: A target detection result is obtained by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[0035] In some embodiments, the first picture may include, but is not limited to, a picture collected by an image or video capture device in an automatic driving system, and/or an intelligent assisted driving system, and/or an intelligent transportation system. The first picture may also include but is not limited to

a picture pre-stored in a database, or collected through other methods.

[0036] In some embodiments, the target neural network model includes, but is not limited to, a convolutional neural network model, a recurrent neural network model, and a combination of one or more neural network models. Specifically, the target neural network model includes, but is not limited to, the neural network model generated based on the multi-scale attention mechanism and the deep separable convolution model.

[0037] In some embodiments, the target feature map includes, but is not limited to, a picture generated after feature information is extracted from the first picture. The value of each pixel in the first picture may include, but is not limited to, a probability of the corresponding position of each pixel in the first picture being a target object. The value of each pixel in the first picture may also include, but is not limited to, the probability of each pixel in the first picture being the lane line pixel or a probability of each pixel in the first picture being an image background pixel.

[0038] It should be noted that the lane line pixel and the image background pixel may include, but are not limited to, two target objects that are mutually exclusive. In other words, when a pixel is not configured to represent a lane line, the pixel is configured to represent an image background.

[0039] In some embodiments, a marked picture may include, but is not limited to, a picture collected by an image or video capture device in an automatic driving system, and/or an intelligent assisted driving system, and/or an intelligent transportation system. The marked picture may also include but is not limited to a picture pre-stored in a database, or collected through other methods.

[0040] In some embodiments, the image post-processing may include, but is not limited to, normalization, image smoothing, image sharpening, image dilation, image erosion, or another one or a combination of image processing methods.

[0041] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0042] Through the present disclosure, a first picture that is planned to be detected is obtained, and a target feature map is obtained by inputting the first picture into a target neural network model. The target neural network model is a model obtained by training a to-be-trained initial neural network model with a set of marked images. Each marked picture includes a marked image background pixel and a marked lane line pixel. The target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel, and the target neutral network model is configured to determine a distribution of the probability in the first picture. A target detection result is obtained by performing image post-processing on the target feature map. The target detection result is configured to indicate a detected lane line in the first picture. The target detection result is configured to indicate the lane line detected in the first picture. Therefore, the technical problem of low detection efficiency of lane lines in related technologies may be solved, and the technical effects of improving the detection efficiency of lane lines, increasing the detection accuracy of lane lines, and reducing detection costs may be achieved.

[0043] In some embodiments, the obtaining the target feature map by inputting the first picture into

the target neural network model includes: obtaining a first feature map by inputting the first picture into a first convolutional layer; obtaining a second feature map by inputting the first feature map into a second convolutional layer, wherein the target neural network model includes the second convolutional layer, and the second convolutional layer is configured to increase a weight of a preset region in the first feature map based on the multi-scale attention mechanism; and obtaining the target feature map by performing a first preset processing on the second feature map, wherein the first preset processing includes an up-sampling operation.

[0044] In some embodiments, the first preset processing may include, but is not limited to, the up-sampling operation, a feature extraction operation, and other processing methods for generating the target feature map from the second feature map.

[0045] In some embodiments, the target neural network model includes the first convolutional layer, and the first convolutional layer is configured to adjust the resolution of the first picture, and to extract the lane line pixel of the first image after the resolution adjustment. The first convolutional layer may also include, but is not limited to, a convolutional neural network for feature extraction, for example, a lightweight convolutional neural network. The lightweight convolutional neural network may perform operations including, but not limited to, a convolution operation with a convolutional kernel in 1×1, 3×3, 5×5, 7×7 and other size. The convolution operation may be achieved with a preset sliding step size. For example, the step size may be 1, 2, etc.

[0046] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0047] In some embodiments, the obtaining the first feature map by inputting the first picture into the first convolutional layer includes: obtaining the first feature map by inputting the first picture with a resolution less than or equal to a first preset resolution into the first convolutional layer.

[0048] In some embodiments, the first preset resolution may be manually set by an operator, and may also set flexibly based on a computational power of the target neural network model or an original resolution of an image or video stream to which the first picture corresponds.

[0049] In some embodiments, the obtaining the first feature map by inputting the first picture with a resolution less than or equal to the first preset resolution may include, but is not limited to, inputting a picture with a lower resolution into the first convolutional layer to achieve the technical effect of improving the efficiency of feature extraction.

[0050] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0051] In some embodiments, the obtaining the first feature map by inputting the first picture into the first convolutional layer includes: obtaining a first sub-feature map by inputting the first picture into a first sub-convolutional layer, wherein the first sub-convolutional layer is configured to perform a convolution operation, and the first convolutional layer includes the first sub-convolutional layer; obtaining a second sub-feature map by inputting the first sub-feature map into a second sub-convolutional

layer, wherein the second sub-convolutional layer is configured to perform a depth separable convolution operation to extract feature information of the first sub-feature map, and the first convolutional layer includes the second sub-convolutional layer; in response to a resolution of the second sub-feature map being greater than a second preset resolution, reducing a resolution of the first sub-feature map by re-inputting the second sub-feature map into the first sub-convolutional layer; in response to the resolution of the second sub-feature map being less than or equal to the second preset resolution, determining the second sub-feature map to be the first feature map.

[0052] In some embodiments, the first sub-convolutional layer is configured to reduce the resolution of a feature map to improve the efficiency of feature extraction, and the second sub-convolutional layer is configured to perform the depth separable convolution operation to extract the feature information in the first sub-feature map. The second preset resolution may be the same as or different from the first preset resolution.

[0053] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0054] In some embodiments, the obtaining the second sub-feature map by inputting the first sub-feature map into the second sub-convolutional layer includes: obtaining the second sub-feature map by extracting the feature information of the first sub-feature map with a 1×n convolutional kernel and by extracting the feature information of the first sub-feature map with an n×1 convolutional kernel, wherein the n is a positive odd number greater than 1.

[0055] In some embodiments, the first convolutional layer may include, but is not limited to, a special combination of a traditional convolutional neural network and a lightweight convolutional neural network. At the bottom of the above target neural network, due to the relatively large resolution of the input image, usually only a smaller convolutional kernel can be used, resulting in a smaller receptive field of the bottom convolutional kernel, while the use of a larger convolutional kernel will lead to excessive computational cost. By adopting the depth separable convolution with the large convolutional kernel in the second sub-convolutional layer in the first convolutional layer mentioned, and by using 1×n and n×1 convolutions instead of n×n convolutions, the computational cost of the neural network model is reduced while increasing the receptive field of the target neural network model. For an n×n convolutional kernel, assuming that the number of input channels is c1 and the number of output channels is c2, the number of parameters of the convolution layer is n×n×c1×c2. When the two convolutions of 1×n and n×1 are used instead, the number of parameters of the convolutional layer is 2×n×c1×c2, and the number of parameters is reduced by (n×n-2×n)×c1×c2. Therefore, the greater the value of n, the more obvious the effect of reducing the number of parameters, while the receptive field remains unchanged.

[0056] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0057] In some embodiments, the obtaining the second feature map by inputting the first feature map into the second convolutional layer includes: obtaining a first type feature map and a plurality of second

type feature maps by inputting the first feature map into the second convolutional layer and by performing a convolution operation on the first feature map with a plurality of convolutional kernels of different sizes included in the second convolution layer; determining a plurality of third type feature maps from the plurality of second type feature maps through a preset statistical method, wherein the plurality of third type feature maps are configured to be performed with a second preset processing such that sizes of the plurality of third type feature maps match each other, the plurality of third type feature maps are configured to be performed with a third preset processing to obtain an attention feature map, and a size of the attention feature map matches a size of the first type feature map; and obtaining the second feature map by performing the third preset processing on the attention feature map and the first type feature map.

[0058] In some embodiments, the plurality of second type feature maps may include, but are not limited to, second type feature maps obtained by performing a convolution operation on the first feature map with a plurality of convolutional kernels of different sizes.

[0059] In some embodiments, the preset statistical method may include, but is not limited to, norm formulas in statistics. Specifically, the preset statistical method may include, but is not limited to, vector norms, matrix norms, etc.

[0060] In some embodiments, the second preset processing may include, but is not limited to, calling a function to adjust the number of channels of a feature map, for example, a reshape operation to adjust the number of rows or columns of a feature vector corresponding to the feature map.

[0061] In some embodiments, the obtaining the second feature map by inputting the first feature map into the second convolutional layer may also include, but is not limited to, increasing a weight of an important region in the first feature map by the multi-scale attention mechanism in a region close to an output layer of the neural network model (for example, the weight of the important region in the first feature map includes but is not limited to a weight of a related region at which a lane line is prone to appear), thereby improving the detection accuracy of the neural network model.

[0062] Specifically, assuming that the resolution of the first feature map is w, h, and the number of channels is c. First, the first feature map is performed with convolution operations with convolutional kernels including but not limited to three convolutional kernels of different scales (1×1, 3×3, and 5×5, corresponding to the aforementioned plurality of convolutional kernels of different sizes) respectively. The use of the different convolutional kernels is to fuse information of lane line elements at different receptive field scales. Three feature maps are output, specifically including two second type feature maps and one first type feature map. While the resolution remains unchanged, and the number of channels is 0.5×c, 0.5×c, and c, respectively. The 3×3 and 5×5 second type feature maps are configured to calculate a correlation between elements to determine the importance of each pixel position in the feature map for global inference. Then element value at each position is counted from channel dimensions using a statistical method. A calculation formula may be as follows.

[0063] $$x=\sum_{i=0}^{c} \mathrm{abs}(x_i)$$

[0064] where x is a counted element value, x_i is element values on different channels at a same location. After calculation, the resolution of the output feature map remains unchanged, and the number of channels becomes 1. Then, operations may include, but are not limited to, processing the output feature map into (w×h)×1 and 1×(w×h) (corresponding to the aforementioned third type feature maps) through the reshape function (corresponding to the second preset processing), obtaining a matrix with a dimension of (w×h)×(w×h) through matrix multiplication (corresponding to the third preset processing) of the (w×h)×1 and 1×(w×h) feature maps, obtaining the attention feature map by performing a softmax operation on the matrix, processing a feature map generated by the 1×1 convolutional kernel (corresponding to the aforementioned first type feature map) into a dimension of c×(w×h) through the reshape function, obtaining another matrix with a dimension of c×(w×h) by performing matrix multiplication of the processed feature map and the attention feature map, processing the another matrix into a c×w×h feature map through the reshape function, and taking the processed another matrix as the second feature map.

[0065] It should be noted that, in middle and top layers of the target neural network, the dimension of the output feature is greatly reduced compared to the bottom of the target neural network. Ordinary convolution is used for feature extraction, and the convolutional kernel is adopted with a larger kernel to continue to maintain the receptive field of the convolutional kernel. 1×n and n×1 convolutions are adopted instead of n×n convolution to reduce the network computational cost.

[0066] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0067] Through the embodiments, the target neural network generated based on the attention mechanism enables the output target feature map to more effectively and accurately reflect each pixel corresponding to the position of the each pixel in the input image, and achieve the detection accuracy of lane lines.

[0068] In some embodiments, before the obtaining the target feature map by inputting the first picture into the target neural network model, the method further includes: obtaining a first sample picture and a label picture corresponding to the first sample picture; and obtaining the target neural network model by training a to-be-trained initial neutral network model with the first sample picture and the label picture, wherein the target neural network model is configured to be trained through a loss function as follows.

[0069] $$loss = C(l_p, l_t) + \alpha\, C(b_p, b_t)$$

[0070] where C represents a cross entropy loss function, l_p represents the lane line pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, l_t represents the lane line pixel marked by the label picture, b_p represents the image background pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, b_t represents the image background pixel marked by the label picture, and α is a preset parameter greater than 0.

[0071] In some embodiments, since the number of the image background pixels is usually far greater than the number of the lane line pixels, the preset α is configured to alleviate the impact of category

imbalance. During the training process, the α may be set to be 0.1.

[0072] In some embodiments, during the training process of the initial neural network model, operations may include, but are not limited to, configuring an initial learning rate to a preset value, for example, 0.01; configuring a learning rate attenuation strategy to a preset strategy, for example, every 10,000 iterations, a learning rate is multiplied by 0.1; and configuring a total number of iterations to another preset value, for example, 60,000 iterations.

[0073] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0074] In some embodiments, the obtaining the target detection result by performing the image post-processing on the target feature map includes: obtaining a lane line result segmentation map by performing a binarization processing on the target feature map; obtaining a processed lane line result segmentation map containing a plurality of connected domains by preprocessing the lane line result segmentation map with an image erosion operation and an image dilation operation; obtaining a target detection result by deleting at least one of the plurality of connected domains that does not meet a preset condition and fitting the remaining of the plurality of connected domains that meets a preset condition to obtain a fitted connected domain, wherein the fitted connected domain included in the target detection result represents a detected lane line in the first picture.

[0075] In some embodiments, the binarization processing may include, but is not limited to, obtaining the lane line result segmentation map by marking an original image based on the target feature map; wherein a part predicted to be the lane line pixel is marked as 1, and another part predicted not to be a lane line, that is, to be the background pixel, is marked as 0.

[0076] It should be noted that FIG. 3 is a schematic view of a lane line detection method based on deep learning according to an embodiment of the present disclosure. As shown in FIG. 3, the lane line pixel may be filtered by performing operations including but not limited to the following operations.

[0077] Operation S1: A width and a height of a circumscribed rectangle of each connected domain 302 are calculated. When the width and height are each less than a corresponding threshold, the connected domain is likely not to be a lane line pixel position, and the connected domain is directly deleted. A scenario before the filtering the lane line pixel may be shown in FIG. 4a, and a result of the filtering the lane line pixel may be shown in FIG. 4b.

[0078] Operation S2: On a processed probability map, each different connected domain may be marked with a different digital id through a Skimage library, representing different lane lines. A result after the processing may be shown in FIG. 4c. The lane line pixels represented by different brightness belong to different lane lines.

[0079] Operation S3: FIG. 5 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure. As shown in FIG. 5, after the connected domains of different ids are obtained, a function y=kx+ b is applied to fit each connected domain separately, where k is the slope and b is the intercept, thereby obtaining a lane line model

corresponding to each connected domain as a candidate lane line 502.

[0080] For example, due to the perspective effect, when parallel lane lines in a three-dimensional space are projected into a two-dimensional space, an angle between each lane line and a horizontal direction will be within a certain range. FIG. 6 is a schematic view of a lane line detection method based on deep learning according to further another embodiment of the present disclosure. As shown in FIG. 6, the range of the angle is required to be determined according to the camera's installation position and focal length parameters. Then, according to the slope of each lane line, the candidate lane lines of which the angle is not within the range are continued to be filtered.

[0081] In some embodiments, the preprocessing of the image erosion operation and the image dilation operation may include but is not limited to one or more operations to obtain the target feature map containing the plurality of connected domains. The preset condition may include but is not limited to a determination based on the size, length and other shape characteristics of the connected domains. For example, by setting a preset length threshold and a preset width threshold, a connected domain is deleted in response to the length of the connected domain being less than the preset length threshold and/or the width of the connected domain being less than the preset width threshold, and the target feature map containing the connected domains that meet the preset condition is retained to determine the final target detection result.

[0082] For example, in some complex road scenes, the lane line features recorded in the target feature map output by the aforementioned target neural network model may show some discontinuities, as shown in FIG. 4a. The scattered and small area lane line prediction pixels may be eliminated using including but not limited to the image erosion operation in OpenCV, because the scattered and small area of lane line prediction pixels are likely to be falsely detected. The prediction pixel area for remaining lane lines is then increased using the image dilation operation in OpenCV, and finally, the connected domains that do not belong to the lane lines are filtered out by configuring the size of the connected domains.

[0083] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0084] In some embodiments, the method further includes:

[0085] re-determining the detected lane line in the first picture by like-for-like merging the fitted connected domains contained in the target detection result by a clustering algorithm.

[0086] In some embodiments, operations may include, but are not limited to, leaving a plurality of lane lines after filtering the non-lane lines. Some candidate lane lines belong to different parts of the same lane line, as shown in the left side of FIG. 5. The two candidate lane lines in the left belong to the same lane line logically, and these lane lines are required to be merged. Clustering is performed using the slope of each candidate lane line. The clustering algorithm used may include, but is not limited to, a mean drift clustering algorithm, which may be implemented by sklearn.cluster.MeanShift function. A clustering radius is required to be determined based on the camera installation location and focal length parameters. The number of clustering centers obtained is the number of lane lines, and the equation of each lane line

after merging is as follows.

[0087] $$y = \frac{1}{n}\sum_{i=1}^{n} k_i x + \frac{1}{n}\sum_{i=0}^{n} b_i$$

[0088] where n indicates the number of candidate lane lines being merged into the lane line, and k_i and b_i indicate the slope and intercept of the i-th candidate lane line respectively.

[0089] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0090] In some embodiments, the target neural network model includes a neural network model including a lightweight convolutional neural network.

[0091] The foregoing is only an exemplary description, and the embodiments do not make any specific limitations.

[0092] The following will further explain the embodiments in combination with specific examples.

[0093] The target neural network model architecture may be adopted with a special combination of a traditional convolutional neural network model and a lightweight convolutional neural network model. At the bottom of the above target neural network model, due to the relatively large resolution of the input image, usually only a smaller convolutional kernel can be used, resulting in a smaller receptive field of the bottom convolutional kernel, while the use of a larger convolutional kernel will lead to excessive computational cost. In the present disclosure, the depth separable convolution of a large convolutional kernel is used at the bottom of the target neural network model, such that the computational cost of the neural network model is reduced while increasing the receptive field of the target neural network model. For an n×n convolutional kernel, assuming that the number of input channels is c1 and the number of output channels is c2, the number of parameters of the convolution layer is n×n×c1×c2. When the two convolutions of 1×n and n×1 are used instead, the number of parameters of the convolutional layer is 2×n×c1×c2, and the number of parameters is reduced by (n×n-2×n)×c1×c2. Therefore, the greater the value of n, the more obvious the effect of reducing the number of parameters, while the receptive field remains unchanged.

[0094] In the middle and top layers of the target neural network model, the dimensionality of the feature map is greatly reduced compared to the bottom of the target neural network model. Ordinary convolution is used for feature extraction, and the convolutional kernel is adopted with a larger kernel to continue to maintain the receptive field of the convolutional kernel. 1×n and n×1 convolutions are adopted instead of n×n convolution to reduce the computational cost of the target neural network model.

[0095] The target neural network model structure is shown in Table 1.

[0096] Table 1

Input	Operation	Convolutional kernel k	Sliding step size s	Number n	Number of output channels c	Convolution kernel void rate
320×184×3	Conv2d,bn,relu	3	2	1	32	-

160×92×32	dewConvSP	5	1	3	64	-
160×92×64	Conv2d,bn,relu	3	2	1	128	-
80×46×128	dewConvSP	5	1	3	128	-
80×46×128	Conv2d,bn,relu	3	2	1	256	-
40×23×256	dewConvSP	5	1	3	512	2
40×23×512	attention	-	1	1	512	-
40×23×512	Tranconv2d,bn,relu	3	2	1	256	-
80×46×256	dewConvSP	5	1	3	64	2
80×46×64	Tranconv2d,bn	3	2	1	2	-
160×92×2	upsample×2	-	-	1	2	-

[0097] Among them, the above dewConvSP convolution operation may include but is not limited to the content shown in Table 2.

[0098] Table 2

Input	Operation	Convolutional kernel k	Sliding step size s	Number of output channels c
w×h×c	dewConv2d,relu	1×k	s	c
w×h×c	dewConv2d,bn,relu	k×1	s	c

[0099] In some embodiments, near the output layer of the target neural network model, a multi-scale attention mechanism is configured to increase the weight of the important region on the feature map (corresponding to the attention convolutional layer in Table 1). To improve the accuracy of target neural network model detection, the calculation flowchart of the proposed attention mechanism is shown in FIG. 7.

[00100] S702: The feature map output by a previous convolutional layer (corresponding to the aforementioned first feature map) is input, assuming that the resolution of the first feature map is w, h, and the number of channels is c.

[00101] S704: A convolution operation is performed on the input image using three convolutional kernels of different scales, wherein the use of the different convolutional kernels is to fuse information of lane line elements at different receptive field scales. Three feature maps are output, the resolution remains unchanged, and the number of channels is 0.5×c, 0.5×c, and c (from top to bottom).

[00102] S706: First two feature maps are configured to calculate the correlation between elements. In order to determine the importance of each location to the global inference. Then element value at each position is counted from channel dimensions using a statistical method. A calculation formula may be as follows.

[00103] $$x=\sum_{i=0}^{c} \mathrm{abs}(x_i)$$

[00104] The foregoing calculation process schematic diagram may include, but is not limited to, as shown in FIG. 8:

[00105] where x is a counted element value, x_i is element values on different channels at a same location. Taking i equal to 4 as an example, after calculation, the resolution of the output feature map remains unchanged, and the number of channels becomes 1. Then the output feature map is processed into (w×h)×1 and 1×(w×h) through the reshape function, a matrix with a dimension of (w×h)×(w×h) is obtained through matrix multiplication, and the attention feature map is obtained by performing a softmax operation on the matrix.

[00106] S708: a feature map generated by H(x) is processed into a dimension of c×(w×h) through the reshape function, another matrix with a dimension of c×(w×h) is obtained by performing matrix multiplication of the processed feature map and the attention feature map, the another matrix is processed into a c×w×h feature map through the reshape function, and the processed another matrix is taken as the second feature map.

[00107] In some embodiments, for all convolution operations in the target neural network model (corresponding to Table 1), as long as the sliding step size is 1, the feature map will be filled with 0 around it before the convolution operation, ensuring that the resolution of the output feature map remains unchanged after the convolution operation. The input of the network is an RGB three-channel color image. The original image is resized to 320×184×3 using a bilinear interpolation, and the label image is resized to 320×184 using a nearest neighbor interpolation algorithm. After a series of convolution processing, the result of lane line segmentation (corresponding to a picture in which each pixel indicates the probability value of whether the location of that pixel is a lane line or not) with the same resolution as the input image. The two channels represent the image background pixel location information and lane line pixel location information, respectively. The loss function used for model training is cross entropy with weights, with the following equation.

[00108] $$loss = C(l_p, l_t) + \alpha\, C(b_p,\ b_t)$$

[00109] where C represents a cross entropy loss function, l_p represents the lane line pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, l_t represents the lane line pixel marked by the label picture, b_p represents the image background pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, b_t represents the image background pixel marked by the label picture, and α is a preset parameter greater than 0.

[00110] It should be noted that since background pixels are usually far more than lane line pixels, α is configured to alleviate the impact of category imbalance, and α may be set to 0.1 during training. The initial learning rate used in training is 0.01. The learning rate decay strategy is that every 10,000 iterations, a learning rate is multiplied by 0.1, and the total number of iterations is set to 60,000.

[00111] After the aforementioned target neural network model outputs the results of the probability map of the lane line segmentation results, the image post-processing and statistical methods are configured to filter the falsely detected lane lines and merge the lane lines. The processing flow chart is shown in FIG. 9. The specific steps may be as follows.

[00112] S902: the target neural network model outputs the target feature map;

[00113] S904: corrosion and dilation processing are performed on the target feature map;

[00114] S906: the connected domains are filtered;

[00115] S908: the connected domains are each marked with an ID;

[00116] S910: the lane lines are fitted according to the connected domains;

[00117] S912: the lane lines are filtered according to the angle range;

[00118] S914: clustering is performed to obtain the final lane line;

[00119] S916: the detection result is output.

[00120] Among them, in some complex road scenes, the lane line features recorded in the target feature map output by the aforementioned target neural network model may show some discontinuities, as shown in FIG. 4a. The scattered and small area lane line prediction pixels may be eliminated using including but not limited to the image erosion operation in OpenCV, because the scattered and small area of lane line prediction pixels are likely to be falsely detected. The prediction pixel area for remaining lane lines is then increased using the image dilation operation in OpenCV, and finally, the connected domains that do not belong to the lane lines are filtered out by configuring the size of the connected domains.

[00121] It should be noted that the above lane line pixels can be filtered by including but not limited to the following methods:

[00122] S1: A width and a height of a circumscribed rectangle of each connected domain 302 as shown in FIG. 3 are calculated. When the width and height are each less than a corresponding threshold, the connected domain is likely not to be a lane line pixel position, and the connected domain is directly deleted. A result after the lane line pixel is processed may be shown in FIG. 4b. The connected domains shown in FIG. 4b do not include part of the connected domains included in FIG. 4a.

[00123] S2: On a processed probability map, each different connected domain may be marked with a different digital id through a Skimage library, representing different lane lines. A result after the processing may be shown in FIG. 4c. The lane line pixels represented by different brightness belong to different lane lines.

[00124] S3: After the connected domains of different ids are obtained, a function $y=kx+b$ is applied to fit each connected domain separately, where k is the slope and b is the intercept, thereby obtaining a lane line model corresponding to each connected domain as a candidate lane line 502 as shown in FIG. 5.

[00125] For example, due to the perspective effect, when parallel lane lines in a three-dimensional space are projected into a two-dimensional space, an angle between each lane line and a horizontal direction will be within a certain range. As shown in FIG. 6, the range of the angle is required to be determined according to the camera's installation position and focal length parameters. Then, according to the slope of each lane line, the candidate lane lines of which the angle is not within the range are continued to be filtered.

[00126] A plurality of lane lines may be left after filtering the non-lane lines. Some candidate lane lines belong to different parts of the same lane line, as shown in the left side of FIG. 5. The two candidate lane lines in the left belong to the same lane line logically, and these lane lines are required to be merged.

Clustering is performed using the slope of each candidate lane line. The clustering algorithm used may include, but is not limited to, a mean drift clustering algorithm, which may be implemented by sklearn.cluster.MeanShift function. A clustering radius is required to be determined based on the camera installation location and focal length parameters. The number of clustering centers obtained is the number of lane lines, and the equation of each lane line after merging is as follows.

[00127] $$y = \frac{1}{n}\sum_{i=1}^{n} k_i x + \frac{1}{n}\sum_{i=0}^{n} b_i$$

[00128] where n indicates the number of candidate lane lines being merged into the lane line, and k_i and b_i indicate the slope and intercept of the i-th candidate lane line respectively.

[00129] With this embodiment, a target neural network model based on a multi-scale attention mechanism is used for the lane line detection task, a lightweight lane line detection network is designed based on the characteristics of each layer of the deep neural network to achieve the use of a lightweight deep neural network in the lane line detection task, which reduces the complexity and computational cost of the network, has advantages when arranged in embedded devices with limited storage capacity and computational power, increases the receptive field of the target neural network model, and alleviates the problem of predicted lane line pixel discontinuity in the lane line segmentation task. Moreover, combined with the output format of the target neural network model, a set of post-processing algorithms may be designed that can accurately detect lane lines based on the coarse extraction of lane line features by the deep neural network, reducing the network's requirement for training datasets. For example, the training can be performed on some general datasets, and the requirement of roughly extracting lane line features can be achieved without specifically collecting the corresponding road scenes for labeling, which reduces the cost of practical application of the algorithm.

[00130] From the description of the above embodiments, it will be clear to those skilled in the art that the method according to the above embodiments can be implemented with the aid of software plus the necessary general purpose hardware platform, or of course by means of hardware, but in many cases the former is the better way of implementation. Based on this understanding, the technical solution of the present disclosure, which essentially or rather contributes to the prior art, may be embodied in the form of a software product, which is stored in a storage medium (e.g. ROM/RAM, disk, CD-ROM) and includes a number of instructions to enable a terminal device (which may be a cell phone, a computer, a server, or a network device, etc.) to perform the method described in various embodiments of the present disclosure.

[00131] The present disclosure also provides a lane line detection apparatus based on deep learning, which is configured to implement the aforementioned embodiments and preferred implementations, and what has been explained will not be repeated. As mentioned below, the term "module" may implement a combination of software and/or hardware with predetermined functions. Although the apparatus described in the following embodiments is preferably implemented by software, hardware or a combination of software and hardware is also possible and conceived.

[00132] FIG. 10 is a structural block view of a lane line detection apparatus based on deep learning

according to an embodiment of the present disclosure. As shown in FIG. 10, the apparatus includes:

[00133] an obtaining module 1002, configured to obtain a first picture that is planned to be detected;

[00134] a first processing module 1004, configured to obtain a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel;

[00135] a second processing module 1006, configured to obtain a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[00136] In some embodiments, the first processing module 1004 includes:

[00137] a first calculation unit, configured to obtain a first feature map by inputting the first picture into a first convolutional layer; wherein the target neural network model includes the first convolutional layer, and the first convolutional layer is configured to adjust a resolution of the first picture and to extract the lane line pixel of the first image after the resolution adjustment;

[00138] a second calculation unit, configured to obtain a second feature map by inputting the first feature map into a second convolutional layer; wherein the target neural network model includes the second convolutional layer, and the second convolutional layer is configured to increase a weight of a preset region in the first feature map based on the multi-scale attention mechanism; and

[00139] a first processing unit, configured to obtain the target feature map by performing a first preset processing on the second feature map; wherein the first preset processing includes an up-sampling operation.

[00140] In some embodiments, the aforementioned apparatus is configured to input the first picture into the first convolutional layer in the following manner to obtain the first feature map: obtaining the first feature map by inputting the first picture with a resolution less than or equal to a first preset resolution into the first convolutional layer.

[00141] In some embodiments, he aforementioned apparatus is configured to input the first picture into the first convolutional layer in the following manner to obtain the first feature map: obtaining a first sub-feature map by inputting the first picture into a first sub-convolutional layer, wherein the first sub-convolutional layer is configured to perform a convolution operation, and the first convolutional layer includes the first sub-convolutional layer; obtaining a second sub-feature map by inputting the first sub-feature map into a second sub-convolutional layer, wherein the second sub-convolutional layer is configured to perform a depth separable convolution operation to extract feature information of the first sub-feature map, and the first convolutional layer includes the second sub-convolutional layer; in response to a resolution of the second sub-feature map being greater than a second preset resolution, reducing a resolution of the first sub-feature map by re-inputting the second sub-feature map into the first sub-convolutional layer; in response to the resolution of the second sub-feature map being less than or

equal to the second preset resolution, determining the second sub-feature map to be the first feature map.

[00142] In some embodiments, the first calculation unit is configured to input the first sub-feature map into the second sub-convolutional layer in the following manner to obtain the second sub-feature map: obtaining the second sub-feature map by extracting the feature information of the first sub-feature map with a 1×n convolutional kernel and by extracting the feature information of the first sub-feature map with an n×1 convolutional kernel, wherein the n is a positive odd number greater than 1.

[00143] In some embodiments, the second calculation unit is configured to input the first feature map into the second convolutional layer in the following manner to obtain the second feature map: obtaining a first type feature map and a plurality of second type feature maps by inputting the first feature map into the second convolutional layer and by performing a convolution operation on the first feature map with a plurality of convolutional kernels of different sizes included in the second convolution layer; determining a plurality of third type feature maps from the plurality of second type feature maps through a preset statistical method, wherein the plurality of third type feature maps are configured to be performed with a second preset processing such that sizes of the plurality of third type feature maps match each other, the plurality of third type feature maps are configured to be performed with a third preset processing to obtain an attention feature map, and a size of the attention feature map matches a size of the first type feature map; and obtaining the second feature map by performing the third preset processing on the attention feature map and the first type feature map.

[00144] In some embodiments, the apparatus is further configured to obtain a first sample picture and a label picture corresponding to the first sample picture; and

[00145] obtain the target neural network model by training a to-be-trained initial neutral network model with the first sample picture and the label picture, wherein the target neural network model is configured to be trained through a loss function as follows.

[00146] $$loss = C(l_p, l_t) + \alpha\ C(b_p,\ b_t)$$

[00147] where C represents a cross entropy loss function, l_p represents the lane line pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, l_t represents the lane line pixel marked by the label picture, b_p represents the image background pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, b_t represents the image background pixel marked by the label picture, and α is a preset parameter greater than 0.

[00148] In some embodiments, the second processing module 1006 includes:

[00149] a second processing unit, configured to obtain a lane line result segmentation map by performing a binarization processing on the target feature map;

[00150] a third processing unit, configured to obtain a processed lane line result segmentation map containing a plurality of connected domains by preprocessing the lane line result segmentation map with an image erosion operation and an image dilation operation;

[00151] a fourth processing unit, configured to obtain a target detection result by deleting at least one of the plurality of connected domains that does not meet a preset condition and fitting the remaining of

the plurality of connected domains that meets a preset condition to obtain a fitted connected domain, wherein the fitted connected domain included in the target detection result represents a detected lane line in the first picture.

[00152] In some embodiments, the apparatus is further configured to:

[00153] re-determine the detected lane line in the first picture by like-for-like merging the fitted connected domains contained in the target detection result by a clustering algorithm.

[00154] In some embodiments, the target neural network model includes a neural network model including a lightweight convolutional neural network.

[00155] It should be noted that each of the above modules can be implemented by software or hardware. For the latter, it can be implemented in the following way, but not limited to: the above modules are all located in the same processor; or, each of the above modules is located in a different processor in any combination.

[00156] An embodiment of the present disclosure also provides a computer-readable storage medium in which a computer program is stored, wherein the computer program is configured to execute the steps in any one of the foregoing method embodiments when executed.

[00157] In this embodiment, the computer-readable storage medium may be configured to store a computer program for executing the following steps:

[00158] S1: A first picture that is planned to be detected is obtained.

[00159] S2. A target feature map is obtained by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel.

[00160] S3: A target detection result is obtained by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[00161] In some embodiments, the foregoing computer-readable storage medium may include, but is not limited to, a USB flash drive, a read-only memory (ROM), a random access memory (RAM), a mobile hard drive, a magnetic disk or an optical disk and other media that can store computer programs.

[00162] An embodiment of the present disclosure also provides an electronic device, including a memory and a processor, the memory stores a computer program, and the processor is configured to run the computer program to execute the steps in any of the foregoing method embodiments.

[00163] In some embodiments, the aforementioned electronic device may further include a transmission device and an input-output device, wherein the transmission device is connected to the processor, and the input-output device is connected to the processor.

[00164] In some embodiments, the processor may be configured to execute the following steps through a computer program:

[00165] S1: A first picture that is planned to be detected is obtained.

[00166] S2. A target feature map is obtained by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel.

[00167] S3: A target detection result is obtained by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

[00168] For specific examples in this embodiment, reference may be made to the examples described in the aforementioned embodiments and exemplary implementations, which will not be repeated here.

[00169] Clearly, it should be understood by those skilled in the art that the modules or steps of the present disclosure described above may be implemented with a generic computing device, they may be centralized on a single computing device or distributed on a network of multiple computing devices, they may be implemented with program code executable by the computing device, thus, they may be stored in a storage device to be executed by the computing device. In some cases, the steps shown or described may be executed in a different order than herein, or they may be implemented separately as individual integrated circuit modules, or multiple modules or steps thereof may be implemented as individual integrated circuit modules. In this way, the present disclosure is not limited to any particular combination of hardware and software.

[00170] The foregoing is only preferred embodiments of the present disclosure and is not intended to limit the present disclosure. To those skilled in the art, the present disclosure is subject to various modifications and variations. Any modification, equivalent replacement, improvement, etc. made within the principles of the present disclosure shall be included in the scope of the present disclosure.

What is claimed is:

1. A lane line detection method based on deep learning, comprising:

obtaining a first picture that is planned to be detected;

obtaining a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model comprises a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel; and

obtaining a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

2. The method according to claim 1, wherein the obtaining the target feature map by inputting the first picture into the target neural network model comprises:

obtaining a first feature map by inputting the first picture into a first convolutional layer;

obtaining a second feature map by inputting the first feature map into a second convolutional layer, wherein the target neural network model comprises the second convolutional layer, and the second convolutional layer is configured to increase a weight of a preset region in the first feature map based on the multi-scale attention mechanism; and

obtaining the target feature map by performing a first preset processing on the second feature map, wherein the first preset processing comprises an up-sampling operation.

3. The method according to claim 2, wherein the obtaining the first feature map by inputting the first picture into the first convolutional layer comprises:

obtaining the first feature map by inputting the first picture with a resolution less than or equal to a first preset resolution into the first convolutional layer.

4. The method according to claim 2, wherein the obtaining the first feature map by inputting the first picture into the first convolutional layer comprises:

obtaining a first sub-feature map by inputting the first picture into a first sub-convolutional layer, wherein the first sub-convolutional layer is configured to perform a convolution operation, and the first convolutional layer comprises the first sub-convolutional layer;

obtaining a second sub-feature map by inputting the first sub-feature map into a second sub-convolutional layer, wherein the second sub-convolutional layer is configured to perform a depth separable convolution operation to extract feature information of the first sub-feature map, and the first convolutional layer comprises the second sub-convolutional layer;

in response to a resolution of the second sub-feature map being greater than a second preset resolution, reducing a resolution of the first sub-feature map by re-inputting the second sub-feature map into the first sub-convolutional layer; and

in response to the resolution of the second sub-feature map being less than or equal to the second preset resolution, determining the second sub-feature map to be the first feature map.

5. The method according to claim 4, wherein the obtaining the second sub-feature map by inputting the first sub-feature map into the second sub-convolutional layer comprises:

obtaining the second sub-feature map by extracting the feature information of the first sub-feature map with a 1×n convolutional kernel and by extracting the feature information of the first sub-feature map with an n×1 convolutional kernel, wherein the n is a positive odd number greater than 1.

6. The method according to claim 2, wherein the obtaining the second feature map by inputting the first feature map into the second convolutional layer comprises:

obtaining a first type feature map and a plurality of second type feature maps by inputting the first feature map into the second convolutional layer and by performing a convolution operation on the first feature map with a plurality of convolutional kernels of different sizes included in the second convolution layer;

determining a plurality of third type feature maps from the plurality of second type feature maps through a preset statistical method, wherein the plurality of third type feature maps are configured to be performed with a second preset processing such that sizes of the plurality of third type feature maps match each other, the plurality of third type feature maps are configured to be performed with a third preset processing to obtain an attention feature map, and a size of the attention feature map matches a size of the first type feature map; and

obtaining the second feature map by performing the third preset processing on the attention feature map and the first type feature map.

7. The method according to claim 1, before the obtaining the target feature map by inputting the first picture into the target neural network model, further comprising:

obtaining a first sample picture and a label picture corresponding to the first sample picture; and

obtaining the target neural network model by training a to-be-trained initial neutral network model with the first sample picture and the label picture, wherein the target neural network model is configured to be trained through a loss function as follows:

$$loss = C(l_p, l_t) + \alpha\ C(b_p,\ b_t)$$

where C represents a cross entropy loss function, l_p represents a lane line pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, l_t represents a lane line pixel marked by the label picture, b_p represents an image background pixel obtained by inputting the first sample picture into the to-be-trained initial neural network model, b_t represents an image background pixel marked by the label picture, and α is a preset parameter greater than 0.

8. The method according to claim 1, wherein the obtaining the target detection result by performing the image post-processing on the target feature map comprises:

obtaining a lane line result segmentation map by performing a binarization processing on the target feature map;

obtaining a processed lane line result segmentation map containing a plurality of connected domains by preprocessing the lane line result segmentation map with an image erosion operation and an image dilation operation; and

obtaining a target detection result by deleting at least one of the plurality of connected domains that does not meet a preset condition and fitting a remaining of the plurality of connected domains that meets a preset condition, wherein a fitted connected domain included in the target detection result represents the detected lane line in the first picture.

9. The method according to claim 8, further comprising:

re-determining the detected lane line in the first picture by like-for-like merging the fitted connected domain included in the target detection result by a clustering algorithm.

10. The method according any one of claims 1-9, wherein the target neural network model comprises a neural network model comprising a lightweight convolutional neural network.

11. A lane line detection apparatus based on deep learning, comprising:

an obtaining module, configured to obtain a first picture that is planned to be detected;

a first processing module, configured to obtain a target feature map by inputting the first picture into a target neural network model; wherein the target neural network model comprises a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel; and

a second processing module, configured to obtain a target detection result by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.

12. A computer-readable storage medium, storing a computer program; wherein the computer program is configured to perform the method according to any one of claims 1-10 when executed.

13. An electronic device, comprising a processor, a memory, and a computer program stored in the memory and executable on the processor; wherein the processor is configured to perform the method according to any one of claims 1-10 when executing the computer program.

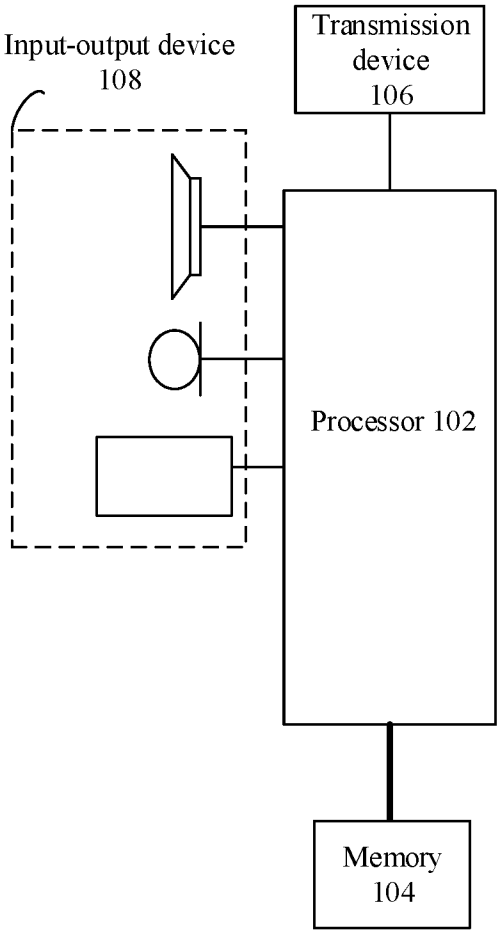

Input-output device
108
Transmission
device
106
Processor 102
Memory
104

FIG. 1

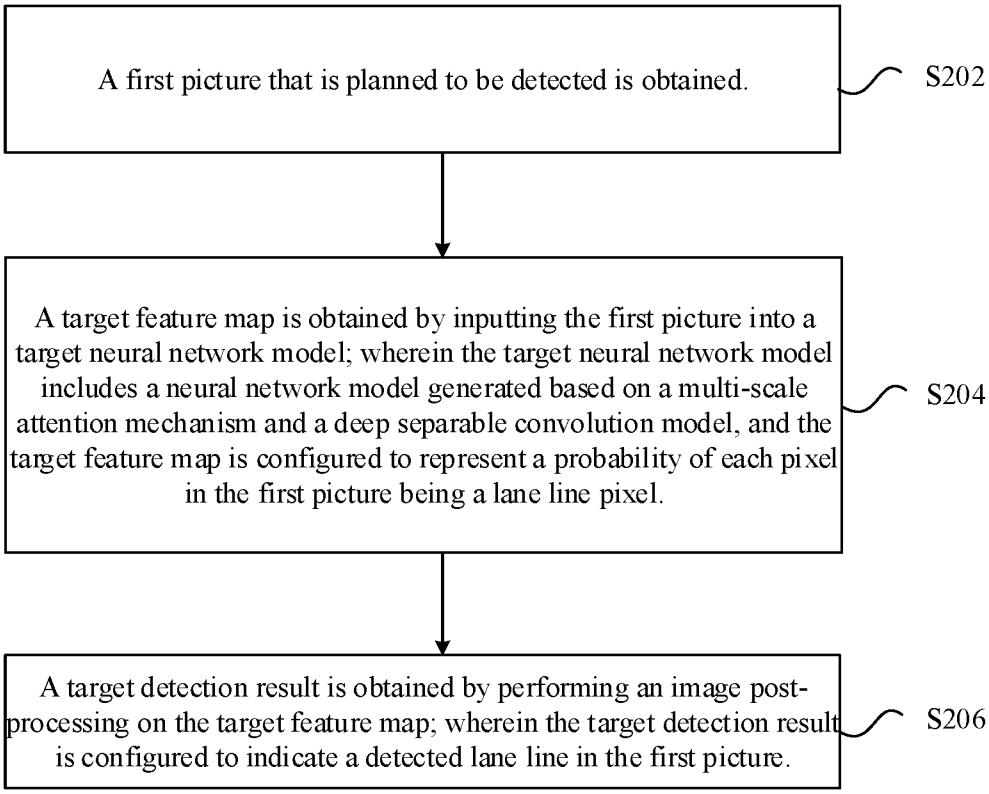

A first picture that is planned to be detected is obtained.
S202
A target feature map is obtained by inputting the first picture into a target neural network model; wherein the target neural network model includes a neural network model generated based on a multi-scale attention mechanism and a deep separable convolution model, and the target feature map is configured to represent a probability of each pixel in the first picture being a lane line pixel.
S204
A target detection result is obtained by performing an image post-processing on the target feature map; wherein the target detection result is configured to indicate a detected lane line in the first picture.
S206

FIG. 2

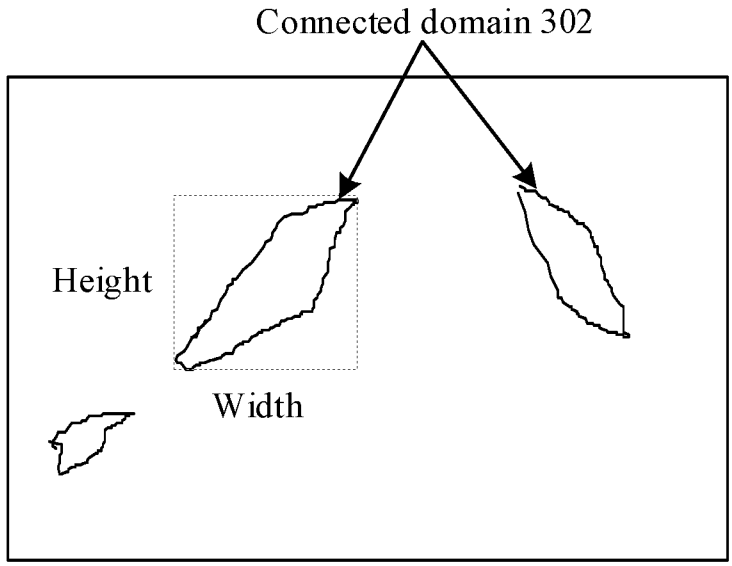

Connected domain 302
Height
Width

FIG. 3

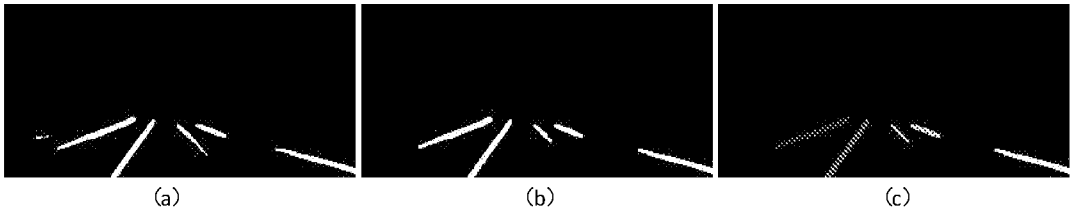


FIG. 4

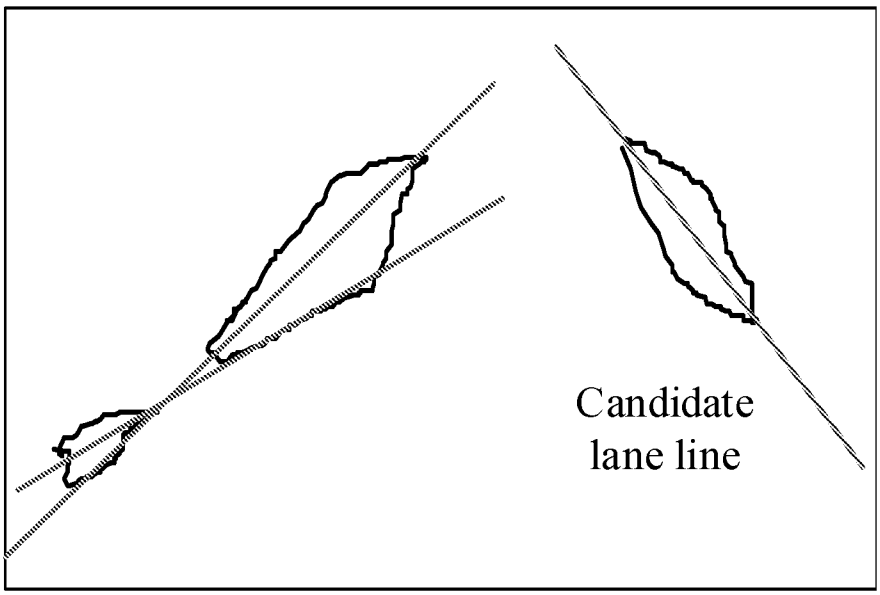


FIG. 5

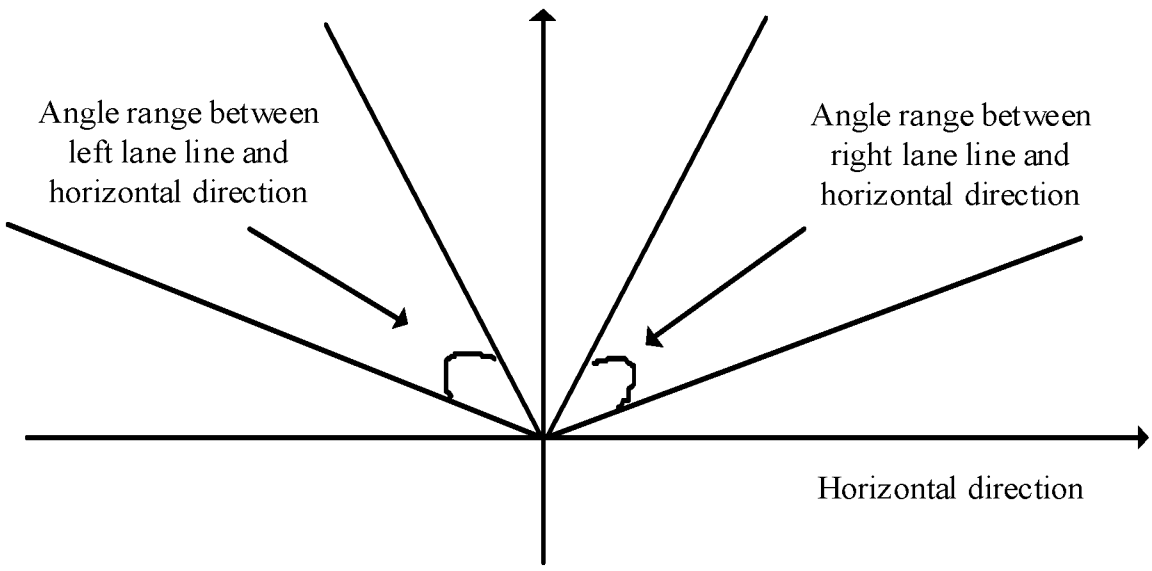


FIG. 6

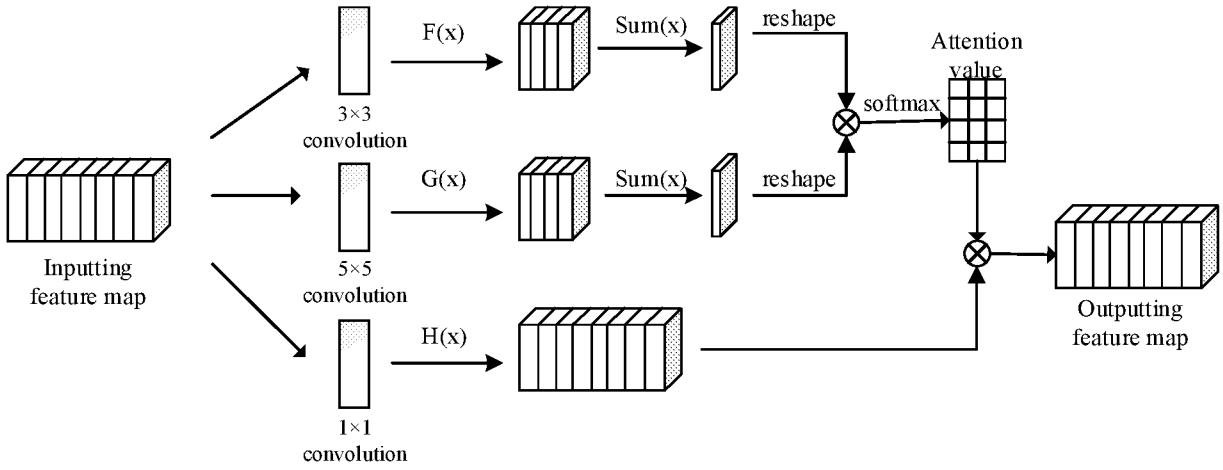

Inputting feature map
3×3 convolution
5×5 convolution
1×1 convolution
F(x)
G(x)
H(x)
Sum(x)
Sum(x)
reshape
reshape
softmax
Attention value
Outputting feature map

FIG. 7

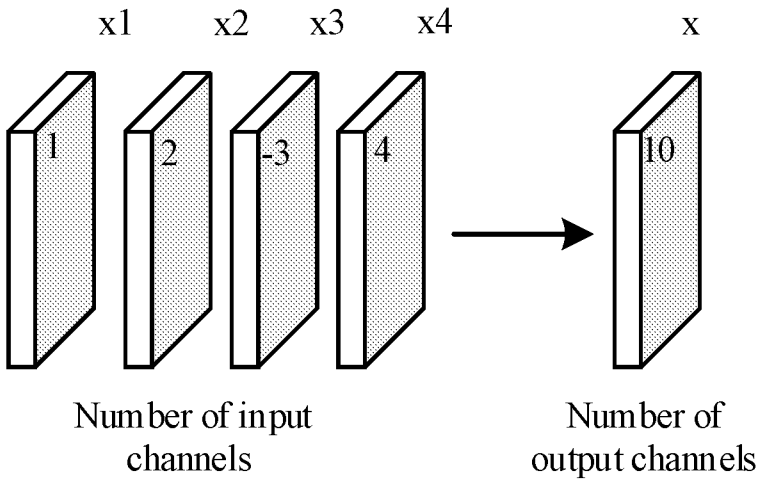

x1
x2
x3
x4
x
1
2
-3
4
10
Number of input channels
Number of output channels

FIG. 8

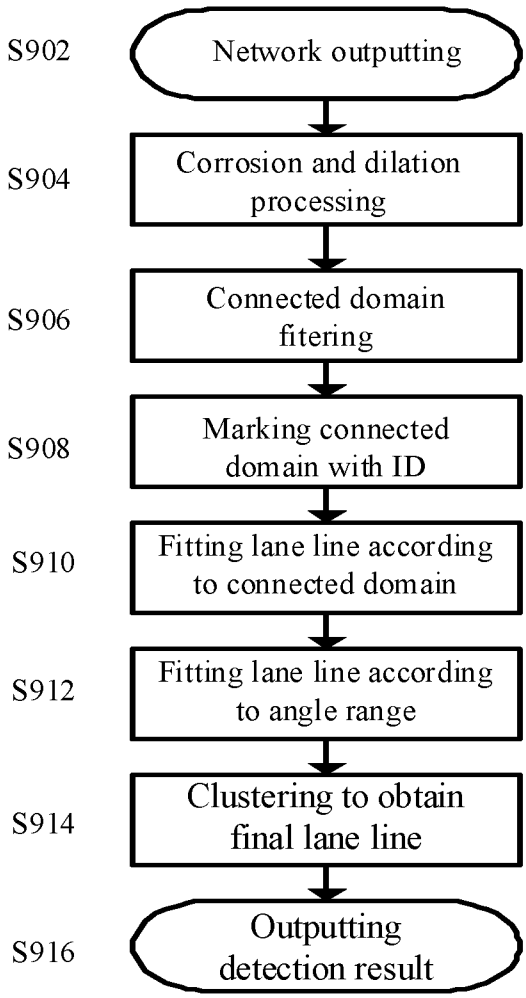

S902
Network outputting
S904
Corrosion and dilation processing
S906
Connected domain fitering
S908
Marking connected domain with ID
S910
Fitting lane line according to connected domain
S912
Fitting lane line according to angle range
S914
Clustering to obtain final lane line
S916
Outputting detection result

FIG. 9

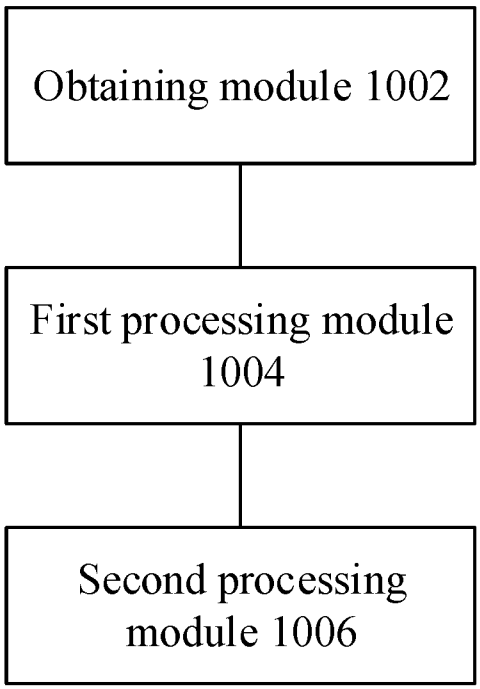

Obtaining module 1002
First processing module 1004
Second processing module 1006

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/132554

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2022.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K; G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, DWPI, CNKI, SIPOABS; lane line, detection, deep learning, picture, image, neural network, attention, deep separable, probability, result

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112287912 A (ZHEJIANG DAHUA TECHNOLOGY CO LTD) 29 January 2021 (2021-01-29) description paragraphs 5-104, figures 1-10	1-13
Y	CN 110276267 A (JIANGSU JINHAIXING NAVIGATION TECHNOLOGY) 24 September 2019 (2019-09-24) description paragraphs 5-63, figures 1-5	1-13
Y	CN 108009524 A (UNIV NORTHWESTERN POLYTECHNICAL) 08 May 2018 (2018-05-08) description paragraphs 5-63, figures 1-5	1-13
A	CN 112036467 A (XUNYIN INTELLIGENT TECHNOLOGY SHANGHAI) 04 December 2020 (2020-12-04) the whole document	1-13

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 February 2022	**01 March 2022**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing** **100088, China** Facsimile No. **(86-10)62019451**	Authorized officer **MENG,Tiange** Telephone No. **86-10-62411653**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2021/132554

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112287912 A	29 January 2021	CN 112287912 B	30 March 2021
CN 110276267 A	24 September 2019	None	
CN 108009524 A	08 May 2018	CN 108009524 B	09 July 2021
CN 112036467 A	04 December 2020	None	