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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
04 June 2020 (04.06.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/108009 A1

(51) International Patent Classification:
G06T 7/00 (2017.01)

(21) International Application Number:
PCT/CN2019/105463

(22) International Filing Date:
11 September 2019 (11.09.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/771,342 26 November 2018 (26.11.2018) US

(71) Applicant: **GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.** [CN/CN]; No.18, Haibin Road, Wusha, Chang'an, Dongguan, Guangdong 523860 (CN).

(72) Inventors: **MENG, Zibo**; 2479 E Bayshore Rd, Suite 110, Palo Alto, California 94303 (US). **HO, Chiuman**; 2479 E Bayshore Rd, Suite 110, Palo Alto, California 94303 (US).

(74) Agent: **SHENZHEN ZHIQUAN INTELLECTUAL PROPERTY OFFICE**; Room 1801, Building 2, Xunmei Technology Plaza, 8 Keyuan Rd, Yuehai Street, 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, 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, SA,

WO 2020/108009 A1

(54) **Title:** METHOD, SYSTEM, AND COMPUTER-READABLE MEDIUM FOR IMPROVING QUALITY OF LOW-LIGHT IMAGES

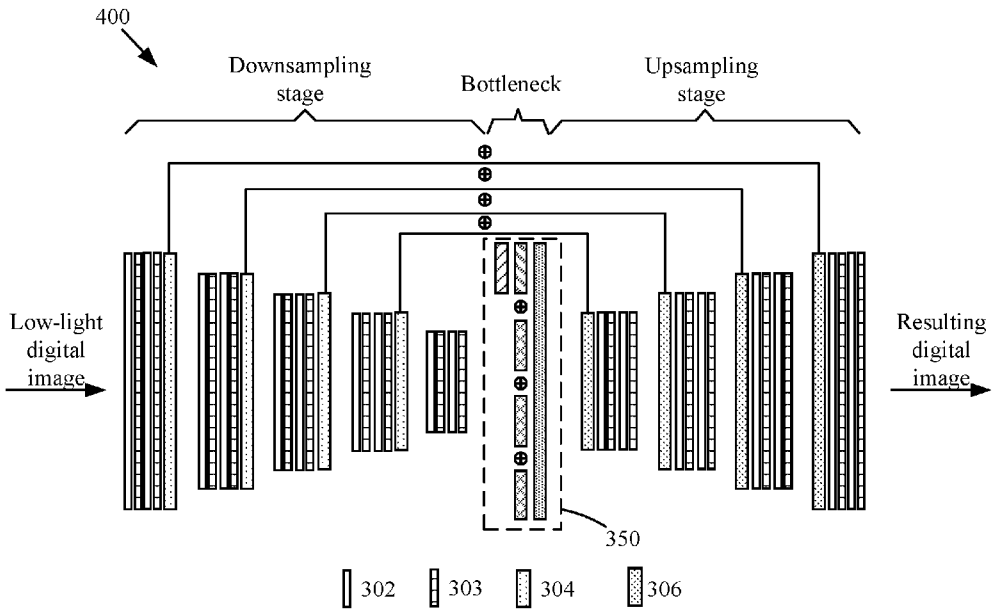

FIG. 6

(57) **Abstract:** In an embodiment, a method includes receiving a low-light digital image; generating, by at least one processor, a resulting digital image by processing the low-light digital image with an encoder-decoder neural network comprising a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the low-light digital image and employed between the downsampling stage and the upsampling stage; and outputting, by the at least one processor, the resulting digital image to an output device. In the network, a channel-wise dropout operation is performed following each of the convolutional layers at the downsampling stage and the upsampling stage to improve generalization ability of the network.

[Continued on next page]

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

Declarations under Rule 4.17:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

METHOD, SYSTEM, AND COMPUTER-READABLE MEDIUM FOR IMPROVING QUALITY OF LOW-LIGHT IMAGES

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to a US application No. 62/771,342 filed on November 26, 2018.

BACKGROUND OF THE DISCLOSURE

1. Field of the Disclosure

[0002] The present disclosure relates to the field of image processing, and more particularly, to a method, system, and computer-readable medium for improving quality of low-light images.

2. Description of the Related Art

[0003] Taking photos having good perceptual quality under low light conditions is extremely challenging due to a low signal-to-noise ratio (SNR). Extending exposure time can acquire visually good images; however, this can easily introduce motion blur, and it is not always applicable in real life.

[0004] To make the low-light images with short exposure time visually plausible, extensive study has been conducted including denoising techniques which aim at removing noises in the images due to the low light condition, and enhancement techniques which are developed for improving the perceptual quality of digital images.

[0005] However, current denoising approaches are generally evaluated using synthetic data, which are not generalized well to real images, and low-light enhancement approaches do not take the noise into consideration. Moreover, since the number of the training dataset is limited, a learning network can easily get overfitted to the training data.

SUMMARY

[0006] An object of the present disclosure is to propose a method, system, and computer-readable medium for improving quality of low-light images.

[0007] In a first aspect of the present disclosure, a method includes:

receiving a digital image;

generating, by at least one processor, a resulting digital image by processing the digital image with an encoder-decoder neural network including a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting, by the at least one processor, the resulting digital image to an output device,

wherein the generating the resulting digital image includes:

performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

[0008] According to an embodiment in conjunction with the first aspect of the present disclosure, in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

[0009] According to an embodiment in conjunction with the first aspect of the present disclosure, in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

[0010] According to an embodiment in conjunction with the first aspect of the present disclosure, before the generating the resulting digital image, the method further includes:

determining whether at least one of a contrast value, a dynamic range, and a signal-to-noise ratio (SNR) of the digital image is lower than a threshold; and

preforming the generating the resulting digital imagein response to determining that at least one of the contrast value, the dynamic range, and the SNR is lower than the threshold.

[0011] According to an embodiment in conjunction with the first aspect of the present disclosure, the generating the resulting digital image further includes:

[0012] concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

[0013] extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

[0014] extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

[0015] According to an embodiment in conjunction with the first aspect of the present disclosure, the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales.

[0016] According to an embodiment in conjunction with the first aspect of the present disclosure, the generating the resulting digital image further includes:

performing a bilinear interpolation operation to the global pooling layer.

[0017] According to an embodiment in conjunction with the first aspect of the present disclosure, the multi-scale context aggregating block includes:

a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

[0018] In a second aspect of the present disclosure, a system includes:

at least one memory configured to store program instructions;

at least one processor configured to execute the program instructions, which cause the at least one processor to perform steps including:

receiving a digital image;

generating a resulting digital image by processing the digital image with an encoder-decoder neural network including a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting the resulting digital image to an output device,

wherein the generating the resulting digital image includes:

performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

[0019] According to an embodiment in conjunction with the second aspect of the present disclosure, in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

[0020] According to an embodiment in conjunction with the first aspect of the present disclosure, in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

[0021] According to an embodiment in conjunction with the second aspect of the present disclosure, the generating the resulting digital image further includes:

concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

[0022] According to an embodiment in conjunction with the second aspect of the present disclosure, the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales, and the multi-scale context aggregating block includes a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

[0023] According to an embodiment in conjunction with the second aspect of the present disclosure, the generating the resulting digital image further includes:

performing a bilinear interpolation operation to the global pooling layer.

[0024] In a third aspect of the present disclosure, a non-transitory computer-readable medium is provided with program instructions stored thereon, that when executed by at least one processor, cause the at least one processor to perform steps including:

receiving a digital image;

generating a resulting digital image by processing the digital image with an encoder-decoder neural network including a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting the resulting digital image to an output device,

wherein the generating the resulting digital image includes:

performing a pooling operation after every few convolutional layers at the downsampling stage to

decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

[0025] According to an embodiment in conjunction with the first aspect of the present disclosure, in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

[0026] According to an embodiment in conjunction with the first aspect of the present disclosure, in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

[0027] According to an embodiment in conjunction with the third aspect of the present disclosure, the generating the resulting digital image further includes:

concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

[0028] According to an embodiment in conjunction with the third aspect of the present disclosure, the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales, and the multi-scale context aggregating block includes a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

[0029] According to an embodiment in conjunction with the third aspect of the present disclosure, the generating the resulting digital image further includes:

performing a bilinear interpolation operation to the global pooling layer.

[0030] In the present disclosure, the digital image is processed using the encoder-decoder neural network. The network includes the convolutional layers classified into the downsampling stage and the upsampling stage, and the multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage. In comparison to existing arts, the present disclosure takes local and global context/color information of the digital image into consideration. Accordingly, the noise can be exhaustively removed and the image can be greatly enhanced for better representation with fruitful details and vivid colors. Moreover, by employing the channel-wise dropout operation, the generalization performance of the network is improved.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] In order to more clearly illustrate the embodiments of the present disclosure or related art, the following figures will be described in the embodiments are briefly introduced. It is obvious that the drawings are merely some embodiments of the present disclosure, a person having ordinary skill in this field can obtain other figures according to these figures without paying the premise.

[0032] FIG. 1 is a diagram illustrating a terminal in accordance with an embodiment of the present disclosure.

[0033] FIG. 2 is a block diagram illustrating software modules and associated hardware of the terminal in accordance with an embodiment of the present disclosure.

[0034] FIG. 3 is a graphical depiction illustrating an encoder-decoder neural network in accordance with an embodiment of the present disclosure.

[0035] FIG. 4 is a graphical depiction showing the U-net architecture of the encoder-decoder neural network depicted in FIG. 3.

[0036] FIG. 5 is a graphic depiction showing the multi-scale context aggregating block depicted in FIG. 3.

[0037] FIG. 6 is a graphical depiction illustrating an encoder-decoder neural network in accordance with another embodiment of the present disclosure.

[0038] FIG. 7A is a diagram illustrating traditional dropout.

[0039] FIG. 7B is a diagram illustrating channel-wise dropout in accordance with the present disclosure.

[0040] FIG. 8 is a flowchart illustrating a method for improving quality of low-light images in accordance with an embodiment of the present disclosure.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0041] Embodiments of the present disclosure are described in detail with the technical matters, structural features, achieved objects, and effects with reference to the accompanying drawings as follows. Specifically, the terminologies in the embodiments of the present disclosure are merely for describing the purpose of the certain embodiment, but not to limit the invention.

[0042] FIG. 1 is a diagram illustrating a terminal 100 in accordance with an embodiment of the present disclosure. Referring to FIG. 1, the terminal 100 includes a camera device 110, a processor module 120, a memory module 130, an output device 140, and a bus 150 connecting to these modules and devices. The terminal 100 has an ability to perform low-light image denoising and enhancement. The terminal 100 can convert low-light images into images with good perceptual quality. The terminal 100 may be implemented by cell phones, smartphones, tablets, notebook computers, desktop computers, or any electronic device having enough computing power to perform the image processing.

[0043] The camera device 110 is configured to capture digital images. When the digital images are captured under low illumination conditions or with an insufficient amount of exposure time, it may be hard to identify the content of the captured digital images. These digital images may have low signal-to-noise ratio (SNR) and are classified as the low-light images. The camera device 110 may be implemented by an RGB camera or a CMYK camera. The camera device 110 is optionally included in the terminal 100. The terminal 100 may perform the image processing to the images with low SNR retrieved from the camera device 110 included in the terminal 100 or any image capturing apparatus outside the terminal 100, or an internal or external storage, or obtained via wired or wireless communication.

[0044] The memory module 130 may be a transitory or non-transitory computer-readable medium that includes a plurality of memory storing program instructions executable by the processor module 120. The processor module 120 includes at least one processor that send signals directly or indirectly to and/or receives signals directly or indirectly from the camera device 110, the memory module 130, and the output device 140

via the bus 150. The processor module 120 is configured to process the digital images (i.e., captured by the camera device 110) with low SNR, by means of a neural network model corresponding to parts of the memory storing program instructions, to generate images with reduced noises and enhanced quality. The neural network model is a key to achieve image denoising and image enhancement in a single process, and will be further described later.

[0045] The images generated by the processor module 120 using the neural network model are outputted by the processor module 120 to the output device 140. The output device 140 may be a storage, a display device, or a wired or wireless communication module for receiving outputted image data from the processor module 120. That is, resulting images with noises reduced and quality enhanced by means of the neural network model may be stored in the storage, displayed on the display device, or transmitted to an external apparatus outside the terminal 10 using an external wired or wireless communication module.

[0046] FIG. 2 is a block diagram illustrating software modules 200 and associated hardware of the terminal 100 in accordance with an embodiment of the present disclosure. The terminal 100 includes the software modules 200 stored in the memory module 130 and executable by the processor module 120. The software modules 200 include a camera control module 202, a low-light image determining module 204, a neural network model 206, and an output control module 208. The camera control module 202 is configured to cause the camera device 110 to take photos to generate a digital image. The low-light image determining module 204 is configured to determine whether the digital image captured by the camera device 110 is a low-light digital image. For example, a contrast value, a dynamic range, and an SNR of the digital image may be used to determine whether it is the low-light digital image. If the contrast value is too low, the dynamic range is too narrow, or the SNR is too small, the digital image is likely to be determined as the low-light digital image. If any one or any combination of the contrast value, the dynamic range, and the SNR is lower than a threshold, the low-light image determining module 204 may classify the captured digital image as the low-light digital image. The low-light digital image is then fed into the neural network model 206 for denoising and enhancement. A resulting digital image is outputted to the output control module 208. The output control module 208 controls transmission of the resulting digital image and decides which device the resulting digital image is to be outputted to, according to a user selection or default settings. The output control module 208 outputs the resulting digital image to the output device 140 such as a display device, a storage, and a wired or wireless communication device.

[0047] FIG. 3 is a graphical depiction illustrating an encoder-decoder neural network 300 in accordance with an embodiment of the present disclosure. The neural network model 206 includes the encoder-decoder neural network 300, as shown in FIG. 3. The low-light digital image is inputted at a left side of the encoder-decoder neural network 300 and the resulting digital image is outputted at a right side of the encoder-decoder neural network 300. Given the low-light digital image, I, the encoder-decoder neural network 300 is employed to learn a mapping, $I'=f(I:w)$, to generate the resulting digital image I' in an end-to-end fashion, where w is a set of learnable parameters of the encoder-decoder neural network 300. Learned parameters and the encoder-decoder neural network 300 are applied to the terminal 100 for image deposing and enhancing. An image taken in a low-light condition with a short exposure is visually unfriendly since it is extremely dark and noisy, where the color and details are invisible to users. By applying the encoder-decoder

neural network 300 and the learned parameters, the image can be enhanced and the noise can be exhaustively removed for better representation on the terminal 100 with fruitful details and vivid colors.

[0048] The pipeline of the encoder-decoder neural network 300 is depicted in FIG. 3. The framework of the encoder-decoder neural network 300 can be divided into two parts, that is, a U-net architecture and a multi-scale context aggregating block 350. FIG. 4 is a graphical depiction showing the U-net architecture of the encoder-decoder neural network 300 depicted in FIG. 3. FIG. 5 is a graphical depiction showing the multi-scale context aggregating block 350 depicted in FIG. 3. The U-net architecture includes a downsampling stage and an upsampling stage, and the multi-scale context aggregating block 350 is employed at a bottleneck between the downsampling stage and the upsampling stage.

[0049] (1) The U-net architecture

[0050] Referring to FIGs. 3 and 4, the U-net architecture includes a plurality of convolutional layers 302 at the downsampling stage and at the upsampling stage. The convolutional layers 302 may be directed to multi-channel feature maps. In an example, each convolutional layer 302 may represent a 3×3 convolutional operation (with a 3×3 filter) and a Leaky ReLU operation. In an example, the U-net architecture may include 18 convolutional layers in total. The resolution gradually decreases and the number of the channels gradually increases for the convolutional layers at the downsampling stage. The resolution gradually increases and the number of the channels gradually decreases for the convolutional layers at the upsampling stage. The low-light digital image firstly goes through downsampling operations to extract abstract features, as well as to reduce the spatial resolution. After the bottleneck, the feature map will go through upscaling operations.

[0051] At the downsampling stage, a pooling layer (e.g., a max pooling layer) 304 is deployed after several convolutional layers 302. For example, the pooling layer 304 is disposed after every two convolutional layers 302. After every few convolutional layers 302, a pooling operation (e.g., a max pooling operation) is performed at the downsampling stage. The pooling operation reduces the resolution of a corresponding feature map. At the upsampling stage, an upscaling layer 306 is deployed before several convolutional layers 302. For example, the upscaling layer 306 is disposed before every two convolutional layers 302. Before every few convolutional layers 302, an upscaling operation is performed at the upsampling stage. The upscaling operation increases the resolution of a corresponding feature map. For example, the upscaling layer 306 is a deconvolutional layer or a transpose convolutional layer.

[0052] Further, the convolutional layers 302 of the downsampling stage and the convolutional layers 302 of the upsampling stage having a (substantially) same resolution (or at substantially same downsampling and upscaling level) with the convolutional layers 302 of the downsampling stage are concatenated. To be described more clearly, the upscaling layer 304 may be formed by upscaling a previous feature map next to the upscaling layer 304 and combining the upscaled feature map with a feature map at the downsampling stage at a level as the same as the upscaled feature map by means of copy and crop operations as needed. The concatenation operation is indicated by a symbol $\oplus$ as depicted in FIG. 4. This operation can effectively preserve the details in an image.

[0053] Examples of the U-net architecture are described in more detail by O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation", in MICCAI, 2015. 4, 5, 7, proposed to segment biomedical images. Other architectures such as an encoder-decoder network having

similar structures throughout an encoder and a decoder are within the contemplated scope of the present disclosure.

[0054] However, the resulting image obtained by only using this U-net architecture to process the low-light digital image may have inconsistent colors at different locations since global context/color information are not taken into consideration. As described below, the present disclosure introduces the global context/color information into the encoder-decoder neural network 300.

[0055] (2) The multi-scale context aggregating block

[0056] Referring to FIGs. 3 and 5, the multi-scale context aggregating block 350 is deployed at the bottleneck between the downsampling stage and the upsampling stage of the encoder-decoder neural network 300. The multi-scale context aggregating block 350 includes a global pooling layer 352 configured to extract global context/color information of the low-light digital image. The global pooling layer 352 may be obtained by means of a pooling operation performed to a previous convolutional layer 302 or a previous feature map next to the global pooling layer 352. The multi-scale context aggregating block 350 also includes a plurality of dilation layers 354 with various dilation rates configured to extract local context/color information of the low-light digital image at different scales. That is, a set of dilated convolutional operations with various dilation rates are employed to extract the local context/color information at different scales. Each dilation layer 354 may be obtained by means of dilation operation performed to a previous convolutional layer 302 or a previous feature map next to a corresponding dilation layer 354. For example, a 2-dilated convolutional operation is performed to a previous feature map to obtain one dilation layer and a 4-dilated convolutional operation is performed to the same to obtain another dilation layer. Dilation operation is an operation to increase the size of receptive field for a feature map, and is a known operation in the art.

[0057] Examples of multi-scale context aggregation are described in more detail by F. Yu, V. Koltun, "Multi-scale context aggregation by dilated convolutions", Proc. Int. Conf. Learn. Representations, 2016, used in image segmentation. Other architectures having similar structures throughout an encoder and a decoder are within the contemplated scope of the present disclosure.

[0058] As depicted in FIG. 5, the global pooling layer 352 and the dilation layers 354 are concatenated. In more details, the global pooling layer 352 and one of the dilation layers 354 are concatenated, and the other dilation layers 354 are concatenated in a cascading fashion with respect to corresponding scales. For example, the global pooling layer 352 and a first dilation layer obtained using a 2-dilated convolutional operation are concatenated, the first dilation layer and a second dilation layer obtained using a 4-dilated convolutional operation are concatenated, the second dilation layer and a third dilation layer obtained using a 8-dilated convolutional operation are concatenated, the third dilation layer and a fourth dilation layer obtained using a 16-dilated convolutional operation are concatenated, and so on.

[0059] The multi-scale context aggregating block 350 further includes a convolutional layer (e.g., a 1×1 convolutional layer) 358 connected after the global pooling layer 352 and the dilation layers 354. In more details, the global pooling layer 352 and the dilation layers 354 are concatenated channel-wisely followed by the convolutional layer 358 to generate a final representation containing multi-scale information of the low-light digital image.

[0060] The inputted low-light digital image may have arbitrary size or resolution, which means that the feature map in the bottleneck has arbitrary size. If a deconvolutional layer is applied after the global pooling

layer 352, the size of kernel in the deconvolutional layer will be dynamic which is almost uncontrollable and not what we want to see. Thus, instead of using the deconvolutional layer, a bilinear upscaling layer 356 is used, in which a bilinear interpolation operation is employed to rescale the feature map back to the same size of the input feature map to perform the concatenation between the global pooling layer 352 and the dilation layers 354 and the following convolutional operations. In more details, the size of the feature map in the global pooling layer 352 is reduced with respect to the feature map in a previous layer. The bilinear interpolation operation can rescale the feature map in the global pooling layer 352 to have a size as the same as the feature map in the previous layer.

[0061] Since the inputted low-light digital image can be of any resolution, the size of the feature maps in the bottleneck of the encoder-decoder neural network 300 depicted in FIG. 3 may still be large. The global context/color information may not be easily observed through the encoder-decoder neural network 300. As a result, the resulting digital image may have inconsistent colors at different locations for a large-sized input image. Moreover, the feature maps in an individual layer of the encoder-decoder neural network 300 may become strongly correlated and this may cause overfitting which may affect the generalization performance of the trained network. To overcome these problems, a channel-wise dropout operation used to improve the generalization ability is employed, which is detailed below.

[0062] FIG. 6 is a graphical depiction illustrating an encoder-decoder neural network 400 in accordance with another embodiment of the present disclosure. In comparison to the embodiment depicted in FIG. 3, the embodiment depicted in FIG. 6 introduces a channel-wise dropout operation following each of the convolutional layers 302 to improve the generalization capability of the network. As illustrated in FIG. 6, a channel-wise dropout layer 303 is deployed following each of the convolutional layers 302 at the downsampling stage and the upsampling stage. The channels or feature maps of the convolutional layers 302 are randomly dropped. More specifically, in the channel-wise dropout operation, each channel or feature map of each of the convolutional layers 302 is given a pre-defined probability to be temporally removed in training the network. That is, all the pixels in a channel or feature map are zeroed out.

[0063] FIG. 7A is a diagram illustrating traditional dropout. FIG. 7B is a diagram illustrating channel-wise dropout in accordance with the present disclosure. In the traditional dropout as depicted in FIG. 7A, a single pixel on the two feature maps is randomly zeroed out. The zeroed-out pixels are indicated by the blocks filled with slash lines. In the channel-wise dropout as depicted in FIG. 7B, a channel or feature map of a convolutional layer is randomly zeroed out, and more particularly, the pixels in a channel or feature map are all zeroed out. For example, the left side of the diagram shown in FIG. 7B indicates a feature map and the right side of the diagram shown in FIG. 7B indicates a zeroed-out feature map of which all the pixels are zeroed out.

[0064] As can be seen from FIGs. 7A and 7B, the traditional dropout operation breaks the spatial correlations in a single feature map, which is of importance for image enhancement. In contrast, the channel-wise dropout operation can keep the spatial correlations, while preventing the feature maps in a single layer from becoming strongly correlated. Accordingly, the embodiment depicted in FIG. 6 can avoid the overfitting and improve the generalization ability of the network.

[0065] Since local and global context/color information is taken into consideration in the present disclosure in low-light image denoising and enhancement, the noise can be exhaustively removed and the image can be greatly enhanced in an end-to-end fashion, leading to better representation with fruitful details and vivid colors.

Moreover, by employing the channel-wise dropout operation, the generalization performance of the network is improved.

[0066] Cost functions

[0067] During the training process, the low-light digital images are fed into the network 300 as input, and a loss function is calculated between the system output and the corresponding long-exposure images. Loss function is a weighted joint loss of ℓ_1 and multi-scale structured similarity index (MS-SSIM), which is defined as follows:

$$\mathcal{L} = \lambda \mathcal{L}^{\ell_1} + (1-\lambda)\mathcal{L}^{MS-SSIM}$$

where λ is set to 0.16 empirically; $\mathcal{L}^{\ell_1}$ is the ℓ_1 loss defined by the following equation:

$$\mathcal{L}^{\ell_1} = \frac{1}{N}\sum_{i \in I}\left|I(i) - \hat{I}(i)\right|$$

where $\hat{I}$ and I are the output image and the ground-truth image, respectively; N is the total number of pixels in the input image.

$\mathcal{L}^{MS-SSIM}$ represents MS-SSIM loss given by the equation below:

$$\mathcal{L}^{MS-SSIM} = 1 - \text{MS} - \text{SSIM}$$

For pixel i, the $\text{MS} - \text{SSIM}$ is defined as:

$$\text{MS} - \text{SSIM}(i) = l_M^{\alpha}(i) \cdot \prod_{j=1}^{M} cs_j^{\beta_j}(i)$$

$$l(i) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$cs(i) = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

Where (x, y) represent the coordinates of pixel i; the means, $i.e. \mu_x . \mu_y$, and standard deviations, $i.e. \sigma_x . \sigma_y$, are calculated with a Gaussian filter, G_g, with zero mean and a standard deviation σ_g; M is the number of levels; and α, β are the weights to adjust the contribution of each component.

[0068] FIG. 8 is a flowchart illustrating a method for improving quality of low-light images in accordance with an embodiment of the present disclosure. Referring to FIGs. 1 to 8, the method includes the following blocks.

[0069] In block 800, the processor module 120 receives a digital image. Preferably, the digital image may be received from the camera device 110 of the terminal 100. In other conditions, the digital image may be transmitted from an external image capturing apparatus, or obtained via wired or wireless communication, or read from an internal or external storage.

[0070] In block 810, the low-light image determining module 204 determines whether the digital image captured by the camera device 110 is a low-light digital image. If any one or any combination of the contrast value, the dynamic range, and the SNR of the digital image is lower than a threshold, the digital image is determined to be the low-light digital image, and go to block 820 to use the encoder-decoder neural network to process the low-light digital image with image denoising and enhancement. If no, the process is terminated.

[0071] In block 820, the encoder-decoder neural network includes a plurality of convolutional layers 302 classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block

350 configured to aggregate multi-scale context information of the low-light digital image and employed between the downsampling stage and the upsampling stage. The encoder-decoder neural network includes a U-net architecture.

[0072] In block 822, in the U-net architecture, a pooling operation is performed after every few convolutional layers 302 at the downsampling stage to decrease spatial resolution and an upscaling operation is performed before every few convolutional layers 302 at the upsampling stage to increase the spatial resolution.

[0073] In block 824, a channel-wise dropout operation is performed following each of the convolutional layers 302 at the downsampling stage and the upsampling stage. In the channel-wise dropout operation, all of pixels in a single channel or feature map of a convolutional layer 302 are zeroed out. More specifically, each channel or feature map of each of the convolutional layers 302 is given a pre-defined probability to be removed.

[0074] In block 826, in the U-net architecture, the convolutional layers 302 of the downsampling stage and the convolutional layers 302 of the upsampling stage having a (substantially) same resolution (or at substantially same downsampling and upscaling level) with the convolutional layers 302 of the downsampling stage are concatenated. The concatenation means feature maps are combined by means of copy and crop operations as needed.

[0075] In block 828, the multi-scale context aggregating block 350 includes the global pooling layer 352, the dilation layers 354, and the convolutional layer (e.g., a 1×1 convolutional layer) 358. The global pooling layer 352 extracts global context/color information of the low-light digital image; and the dilation layers 354 with various dilation rates extract local context/color information of the low-light digital image at different scales. The global pooling layer 352 and one of the dilation layers 354 are concatenated, and the other dilation layers 354 are concatenated in a cascading fashion with respect to corresponding scales. The convolutional layer 358 is connected after the global pooling layer 352 and the dilation layers 354 to generate a final representation containing multi-scale information of the low-light digital image. A bilinear interpolation operation may be performed to the global pooling layer 352 to rescale the size of the feature map in the global pooling layer 352 to a size as (substantially) the same as the input feature map.

[0076] Other details of the encoder-decoder neural network are referred to related descriptions in above context, and are not repeated herein.

[0077] In block 830, the encoder-decoder neural network outputs a resulting digital image and the processor module 120 outputs the resulting digital image to the output device 140 such as a display device, a storage, and a wired or wireless communication device.

[0078] Other details of the method for improving quality of low-light images are referred to related descriptions in above context, and are not repeated herein.

[0079] In the present disclosure, the digital image is processed using the encoder-decoder neural network. The network includes the convolutional layers classified into the downsampling stage and the upsampling stage, and the multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage. In comparison to existing arts, the present disclosure takes local and global context/color information of the digital image into consideration. Accordingly, the noise can be exhaustively removed and the image can be greatly enhanced for

better representation with fruitful details and vivid colors. Moreover, by employing the channel-wise dropout operation, the generalization performance of the network is improved.

[0080] A person having ordinary skill in the art understands that each of the units, modules, algorithm, and steps described and disclosed in the embodiments of the present disclosure are realized using electronic hardware or combinations of software for computers and electronic hardware. Whether the functions run in hardware or software depends on the condition of application and design requirement for a technical plan. A person having ordinary skill in the art can use different ways to realize the function for each specific application while such realizations should not go beyond the scope of the present disclosure.

[0081] It is understood by a person having ordinary skill in the art that he/she can refer to the working processes of the system, device, and module in the above-mentioned embodiment since the working processes of the above-mentioned system, device, and module are basically the same. For easy description and simplicity, these working processes will not be detailed.

[0082] It is understood that the disclosed system, device, and method in the embodiments of the present disclosure can be realized with other ways. The above-mentioned embodiments are exemplary only. The division of the modules is merely based on logical functions while other divisions exist in realization. It is possible that a plurality of modules or components are combined or integrated in another system. It is also possible that some characteristics are omitted or skipped. On the other hand, the displayed or discussed mutual coupling, direct coupling, or communicative coupling operate through some ports, devices, or modules whether indirectly or communicatively by ways of electrical, mechanical, or other kinds of forms.

[0083] The modules as separating components for explanation are or are not physically separated. The modules for display are or are not physical modules, that is, located in one place or distributed on a plurality of network modules. Some or all of the modules are used according to the purposes of the embodiments.

[0084] Moreover, each of the functional modules in each of the embodiments can be integrated in one processing module, physically independent, or integrated in one processing module with two or more than two modules.

[0085] If the software function module is realized and used and sold as a product, it can be stored in a readable storage medium in a computer. Based on this understanding, the technical plan proposed by the present disclosure can be essentially or partially realized as the form of a software product. Or, one part of the technical plan beneficial to the conventional technology can be realized as the form of a software product. The software product in the computer is stored in a storage medium, including a plurality of commands for a computational device (such as a personal computer, a server, or a network device) to run all or some of the steps disclosed by the embodiments of the present disclosure. The storage medium includes a USB disk, a mobile hard disk, a read-only memory (ROM), a random access memory (RAM), a floppy disk, or other kinds of media capable of storing program codes.

[0086] While the present disclosure has been described in connection with what is considered the most practical and preferred embodiments, it is understood that the present disclosure is not limited to the disclosed embodiments but is intended to cover various arrangements made without departing from the scope of the broadest interpretation of the appended claims.

What is claimed is:

1. A method, comprising:

receiving a digital image;

generating, by at least one processor, a resulting digital image by processing the digital image with an encoder-decoder neural network comprising a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting, by the at least one processor, the resulting digital image to an output device,

wherein the generating the resulting digital image comprises:

performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

2. The method according to claim 1, wherein in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

3. The method according to claim 1, wherein in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

4. The method according to claim 1, wherein before the generating the resulting digital image, the method further comprises:

determining whether at least one of a contrast value, a dynamic range, and a signal-to-noise ratio (SNR) of the digital image is lower than a threshold; and

performing the generating the resulting digital image in response to determining that at least one of the contrast value, the dynamic range, and the SNR is lower than the threshold.

5. The method according to claim 1, wherein the generating the resulting digital image further comprises:

concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

6. The method according to claim 5, wherein the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales.

7. The method according to claim 6, wherein the generating the resulting digital image further comprises:

performing a bilinear interpolation operation to the global pooling layer.

8. The method according to claim 7, wherein the multi-scale context aggregating block comprises:

a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

9. A system, comprising:

at least one memory configured to store program instructions;

at least one processor configured to execute the program instructions, which cause the at least one processor to perform steps comprising:

receiving a digital image;

generating a resulting digital image by processing the digital image with an encoder-decoder neural network comprising a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting the resulting digital image to an output device,

wherein the generating the resulting digital image comprises:

performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

10. The system according to claim 9, wherein in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

11. The system according to claim 9, wherein in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

12. The system according to claim 9, wherein the generating the resulting digital image further comprises:

concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

13. The system according to claim 12, wherein the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales, and the multi-scale context aggregating block comprises a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

14. The system according to claim 13, wherein the generating the resulting digital image further comprises:

performing a bilinear interpolation operation to the global pooling layer.

15. A non-transitory computer-readable medium with program instructions stored thereon, that when executed by at least one processor, cause the at least one processor to perform steps comprising:

receiving a digital image;

generating a resulting digital image by processing the digital image with an encoder-decoder neural network comprising a plurality of convolutional layers classified into a downsampling stage and an upsampling stage, and a multi-scale context aggregating block configured to aggregate multi-scale context information of the digital image and employed between the downsampling stage and the upsampling stage; and

outputting the resulting digital image to an output device,

wherein the generating the resulting digital image comprises:

performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution;

performing an upscaling operation before every few convolutional layers at the upsampling stage to increase the spatial resolution; and

performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.

16. The non-transitory computer-readable medium according to claim 15, wherein in the performing the channel-wise dropout operation, each channel or feature map of each of the convolutional layers is given a pre-defined probability to be removed.

17. The non-transitory computer-readable medium according to claim 15, wherein in the performing the channel-wise dropout operation, all of pixels in a single channel or feature map of each of the convolutional layer are zeroed out.

18. The non-transitory computer-readable medium according to claim 15, wherein the generating the resulting digital image further comprises:

concatenating the convolutional layers of the downsampling stage and the convolutional layers of the upsampling stage having a same resolution with the convolutional layers of the downsampling stage;

extracting, by a global pooling layer of the multi-scale context aggregating block, global context information of the digital image; and

extracting, by a plurality of dilation layers with various dilation rates of the multi-scale context aggregation block, context information of the digital image at different scales.

19. The non-transitory computer-readable medium according to claim 18, wherein the global pooling layer and one of the dilation layers are concatenated, and the other dilation layers are concatenated in a cascading fashion with respect to corresponding scales, and the multi-scale context aggregating block comprises a 1×1 convolutional layer connected after the global pooling layer and the dilation layers.

20. The non-transitory computer-readable medium according to claim 19, wherein the generating the resulting digital image further comprises:

performing a bilinear interpolation operation to the global pooling layer.

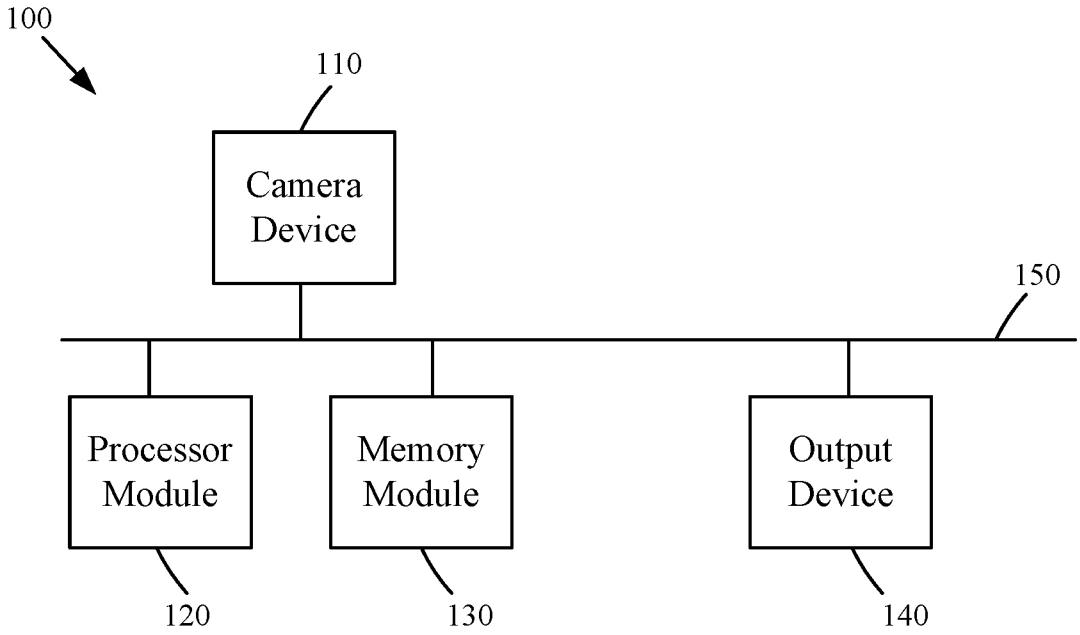
100
110
Camera Device
150
Processor Module
Memory Module
Output Device
120
130
140

FIG. 1

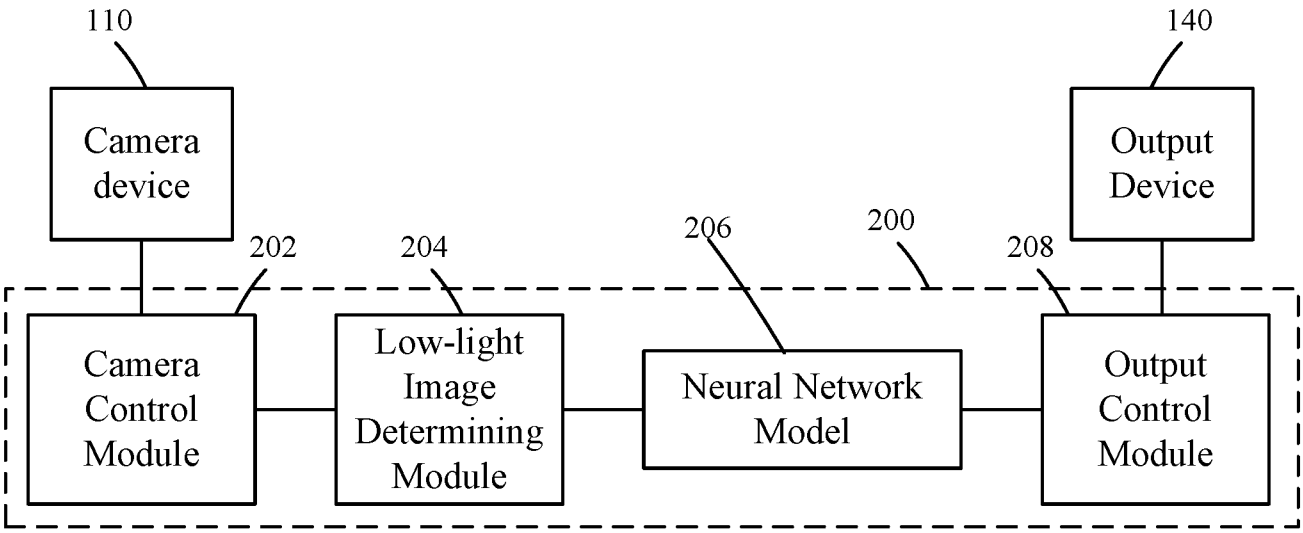
110
Camera device
140
Output Device
202
204
206
200
208
Camera Control Module
Low-light Image Determining Module
Neural Network Model
Output Control Module

FIG. 2

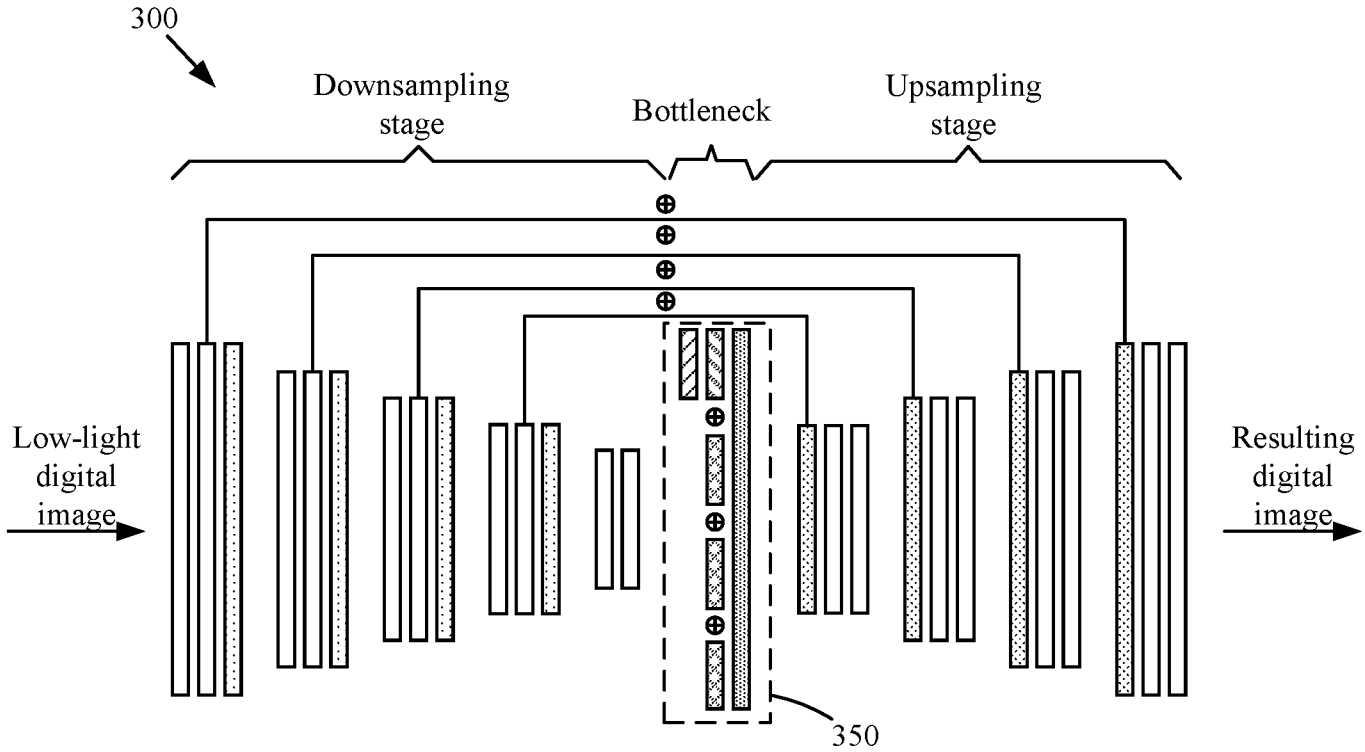
300
Downsampling stage
Bottleneck
Upsampling stage
Low-light digital image
Resulting digital image
350

FIG. 3

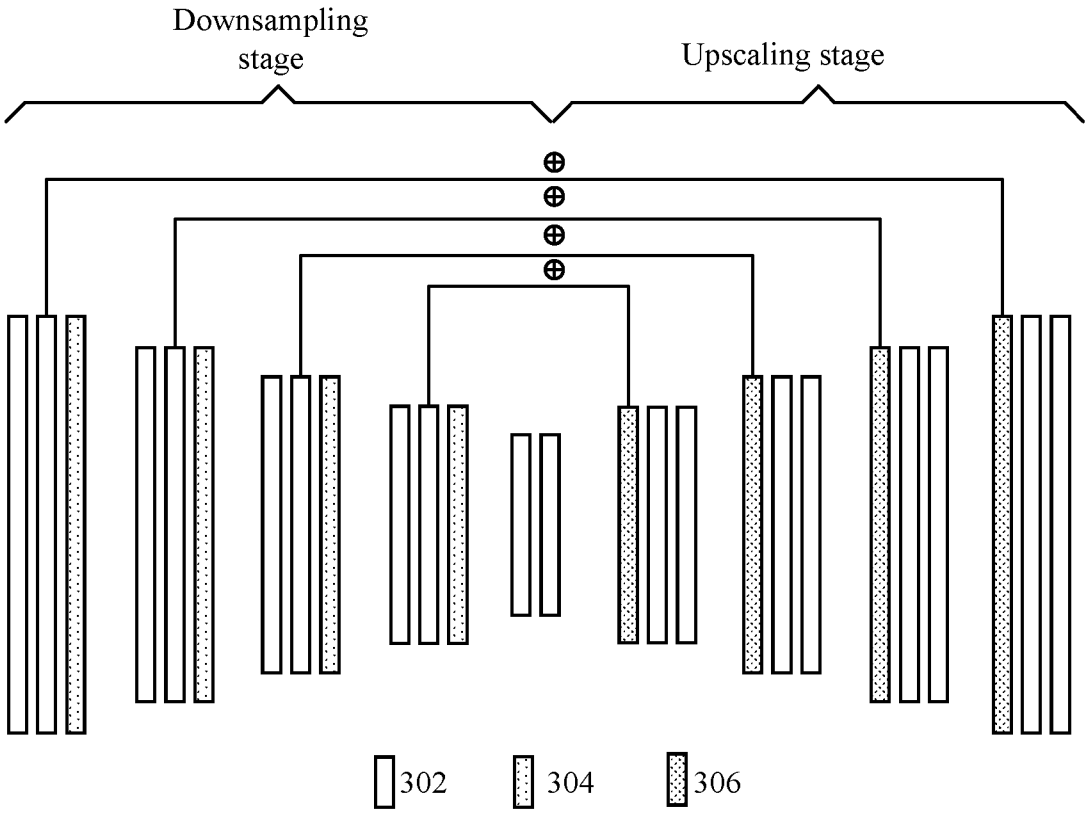
Downsampling stage
Upscaling stage
302
304
306

FIG. 4

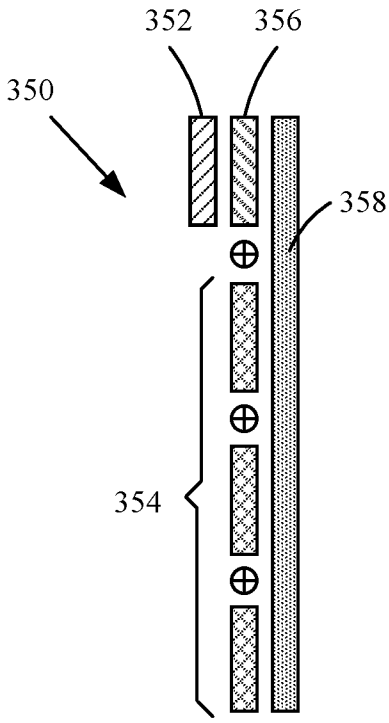
352
356
350
358
354

FIG. 5

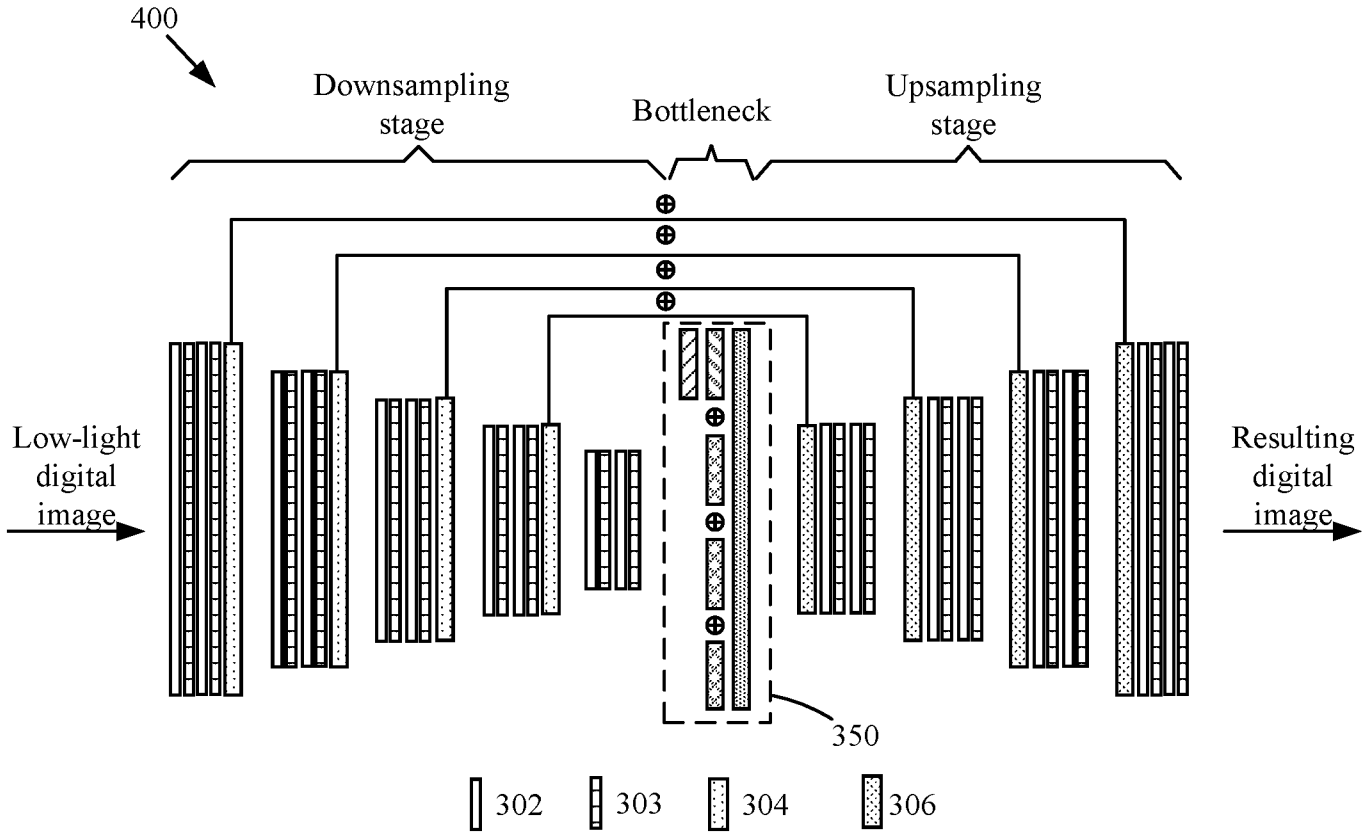
400
Downsampling stage
Bottleneck
Upsampling stage
Low-light digital image
Resulting digital image
350
302
303
304
306

FIG. 6

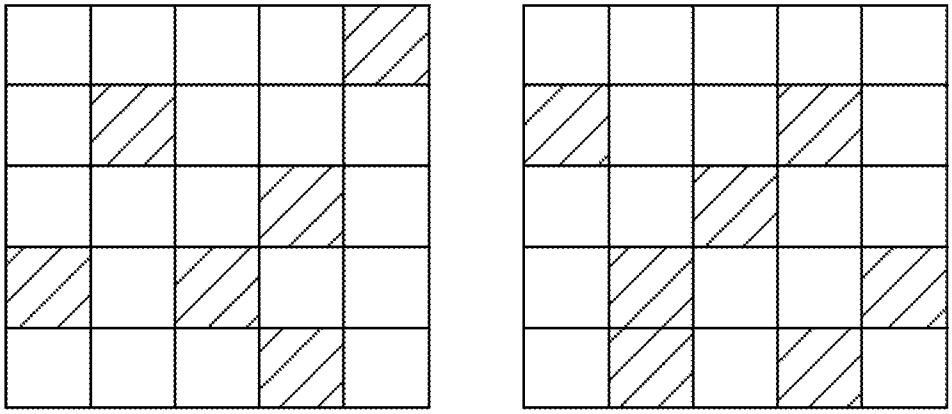

FIG. 7A

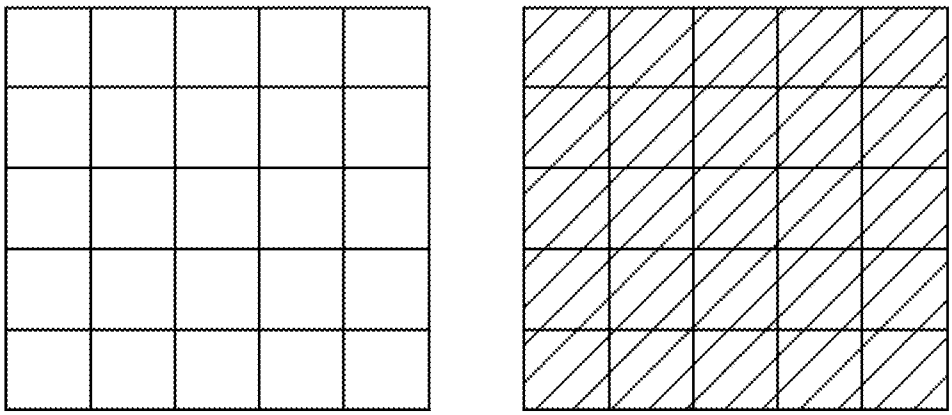

FIG. 7B

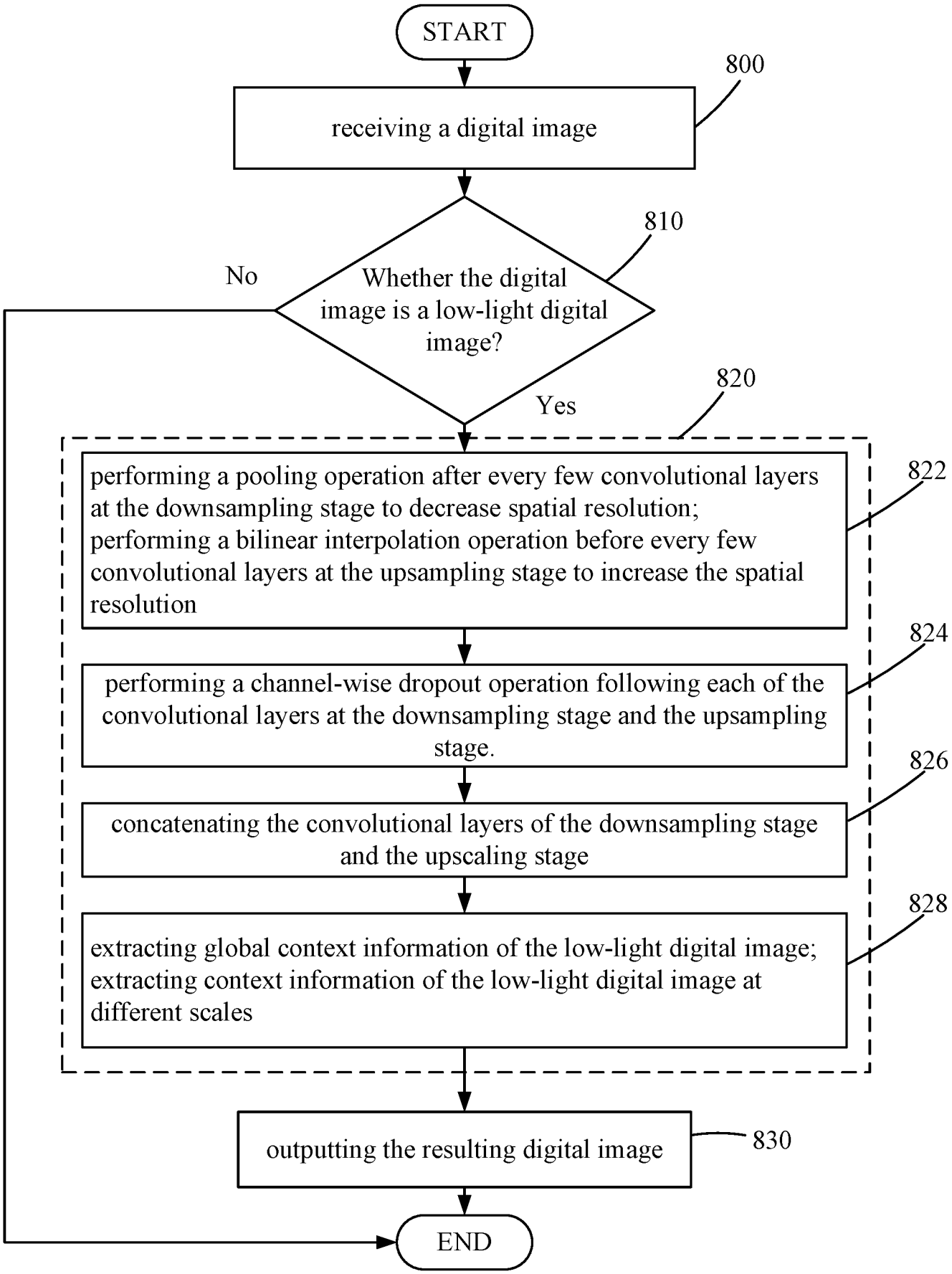
START
800
receiving a digital image
810
Whether the digital image is a low-light digital image?
No
Yes
820
performing a pooling operation after every few convolutional layers at the downsampling stage to decrease spatial resolution; performing a bilinear interpolation operation before every few convolutional layers at the upsampling stage to increase the spatial resolution
822
performing a channel-wise dropout operation following each of the convolutional layers at the downsampling stage and the upsampling stage.
824
concatenating the convolutional layers of the downsampling stage and the upscaling stage
826
extracting global context information of the low-light digital image; extracting context information of the low-light digital image at different scales
828
outputting the resulting digital image
830
END

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/105463

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(2017.01)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)

G06T

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)

CNPAT,CNKI,WPI,EPODOC,IEEE: image, neural, network, convolutional, layer?, down, up, sampl+, downsampl+, upsampl+, context, pool+, upscal+, dropout, concatenat+, u-net, enhanc+, low, light, dilat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107016665 A (UNIV.ZHEJIANG) 04 August 2017 (2017-08-04) description, paragraphs [0057]-[0058], figure 2	1-20
A	CN 107967514 A (HTC CORP.) 27 April 2018 (2018-04-27) the whole document	1-20
A	CN 108596915 A (SHENZHEN FUTURE MEDIA TECHNOLOGY INST. ET AL.) 28 September 2018 (2018-09-28) the whole document	1-20
A	CN 108846835 A (UNIV. XIDIAN) 20 November 2018 (2018-11-20) the whole document	1-20
A	CN 108805874 A (THE THIRD RESEARCH INSTITUTE OF CHINA ELECTRONICS TECH. GROUP CORPORATION) 13 November 2018 (2018-11-13) the whole document	1-20
A	WO 2018125580 A1 (KONICA MINOLTA LABORATORY U.S.A., INC.) 05 July 2018 (2018-07-05) the whole document	1-20
A	WO 2018140596 A2 (ARTERYS INC.) 02 August 2018 (2018-08-02) the whole document	1-20

☐ 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
06 November 2019	**27 November 2019**
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 **WANG,Conglei** Telephone No. **86-(10)-53961717**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/105463

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 107016665 A	04 August 2017	None	
CN 107967514 A	27 April 2018	US 2018114294 A1	26 April 2018
		TW 201816716 A	01 May 2018
CN 108596915 A	28 September 2018	None	
CN 108846835 A	20 November 2018	None	
CN 108805874 A	13 November 2018	None	
WO 2018125580 A1	05 July 2018	US 2019205758 A1	04 July 2019
WO 2018140596 A2	02 August 2018	US 2018218497 A1	02 August 2018