A DEEP LEARNING MODEL OF DEFECT DETECTION METHOD AND SYSTEM FOR PLASTIC INJECTION MOLDING PRODUCTS

FIELD OF THE INVENTION

[0001] The present invention relates to a method and system for detecting defects for plastic injection molding products, in particular a defect detection method and system based on deep learning model for plastic injection molding products.

BACKGROUND OF THE INVENTION

[0002] In injection molding industry, products are produced under massive and fast production. Since the injection molding machine industry can adjust molds to meet various companies' needs, the products are widely applied in many fields, such as pharmaceutical, telecommunications, transportation, consumer products and medical sectors. During the manufacturing process, the variation of the injection molding machine condition may cause defective production. For example, inadequate injection pressure may lead to product with sink mark or short shot in which the said short shot is a structural defect. A tiny defect within the injection molding product may lead to a disastrous accident. To avoid the production of defective products, the common practice in the injection molding industry is to conduct a defect detection after product production.

[0003] Traditional defect detection methods rely on manual detection, which is time-consuming and expensive. Furthermore, after extended periods of operation, efficiency decreases while the rate of human error increases. However, most of the defect detection methods in the market are based on image classification with single product with mono-class prediction. When more than one product or a product with more than one defect is detected using the said traditional methods, a defective region cannot be clearly identified.

[0004] In recent days, there have been several approaches for defect detection on plastic injection molding products. Image classification and object detection are common approaches for defect detection on plastic injection molding products. The lack of information about the defective region is the major limitation of image classification.

Image classification can only predict a class for the input frame without identifying the defective region. Labelling of defective regions for prediction is important for industrial applications since the workers need to know the reason why a product is predicted as defective. Another limitation of image classification is the weakness of the detection of multiple defective classes or multiple objects since only a single class can be predicted for the input frame. In fact, plastic injection molding products are not constrained by only zero or one defect type. The limitation on multi-object detection may lower the detection rate since the products need to be queued and detected one by one. Even worse, the multiple objects that occur accidentally on the frame may cause false classification.

[0005] Apart from that, machine learning has been applied to defect detection on plastic injection molding products these days. The general approach is to perform image classification on the frame of the video stream. If the computational power is not strong enough, the frames per second (FPS) of the detection may be reduced. As a result, the product's detection rate will be reduced directly. Object detection can be achieved by using machine learning with the YOLOv3 structure. Deep learning models with object detection can be classified into two categories: One-stage detectors and two-stage detectors. In one-stage detectors, the bounding box and corresponding class are predicted simultaneously, whereas in two-stage detectors, classification is followed by the previously predicted bounding box. One-stage detectors detect faster with less procedures, whereas two-stage detectors predict more accurately at the expense of speed. YOLOv3 is a one-stage detector, hence it is fast enough for real-time detection. However, the accuracy of class prediction and localization is lower when compared to two-stage detectors. Therefore, limitations in localization accuracy still exist in recent works.

[0006] China Patent Publication No. 104850858 B discloses a kind of injection-molded item defects detection recognition method. The method comprises the following steps: normal image with the injection-molded item that there is known defect of collection, image is classified and generates sample; multilayer convolutional neural networks model is built; training the convolutional neural network with the sample images; and the trained convolutional neural networks model can classify and identify the image of measured injection-molded item for defect detection. Iterations epoch of convolutional neural networks model is set in the method, wherein the convolutional neural networks model training is complete when rear stops iteration. However, the method may require high network parameters, which will decrease the speed of the detection. Hence, there is a need to provide a method and system of defect detection that can boost the speed of the algorithm and/or the detection system.

[0007] China Patent Publication No. 110136116 A discloses a kind of injection molding pump defect inspection method, device, equipment and storage medium. The method comprises the steps of: obtaining injection molding pump sample image and extracting the coordinate of the rejected region marked in advance in the injection molding pump sample image, generating corresponding defect text information; inputting the injection molding pump sample image and defect text information in a deep learning neural network, training the deep learning training neural network to obtain target neural network; validating the target neural network using injection molding pump test image, adjusting the network parameter of target nerve network to obtain the defects detection model; and detecting defects. However, the utilization of model YOLOv3 in the method may become limited when detecting multiple products and/or multiple classes. Said limitation on multiple products detection may reduce the detection rate since the products need to be queued and detected one by one. Even worse, the multiple products that occur accidentally on the frame may cause false classification. Also, the input of image is passed to two convolution layers followed by a bottleneck layer as the beginning of model YOLOv3, which requires more learning parameters and hence reduce the speed of the algorithm. In addition, Feature Pyramid Network (FPN) is applied in YOLOv3 for detecting small objects in a large-scale detector. Poor fine-grained features are used for detection in small-scale detectors and hence the accuracy in small-scale object detection is limited. In YOLOv3, initial anchor boxes are obtained by running the k-means clustering on the dataset, while bounding boxes are predicted from those pre-defined anchor boxes with shifting and scaling. However, the pre-defined anchor boxes may have a high error rate if the aspect ratio of the object in the dataset is highly variable, while the prediction of the bounding box afterwards will be affected. An improved product detection method and system is needed to overcome the shortcomings.

[0008] China Patent Publication No. 111487250 A discloses an intelligent visual detection method and system applied to injection molding defective product detection. The method comprises the steps of: acquiring video stream data of a sample to be detected; decomposing video stream data into single frame image data; detecting the defect of the defective product through a defective product detection deep learning algorithm; classifying the defective products through a defective product classification algorithm; and counting the number of the detected samples and the number of the defective products by an injection part tracking detection algorithm, and displaying the detection result. The method can be used to rapidly and accurately detect products of varying degrees of defects in real time, and has higher robustness and generalization capability compared with the traditional detection method. However, the utilization of model YOLOv3 in the method may become limited when detecting multiple products and/or multiple classes. Such limitation on multiple products detection may reduce the detection rate since the products need to be queued and detected one by one. Even worse, the multiple products occur that accidentally on the frame may cause false classification. Also, the input of image is passed to two convolution layers following by a bottleneck layer as the beginning of model YOLOv3, which requires more learning parameters and hence reduces the speed of the algorithm. In addition, FPN is applied in YOLOv3 for detecting small objects in large-scale detector. Poor fine-grained features are used for detection in small-scale detectors and hence the accuracy in small-scale object detection is limited. In YOLOv3, initial anchors boxes are obtained by running the k-means clustering on the dataset while bounding boxes are predicted from those pre-defined anchor boxes with shifting and scaling. However, the pre-defined anchor boxes may have a high error rate if the aspect ratio of the object in the dataset is highly variable, while the prediction of the bounding box afterwards will be affected. An improved product detection method and system are needed to overcome the shortcomings.

SUMMARY OF THE INVENTION

[0009] It is an objective of the present invention to provide a defect detection method and system with a deep learning model for plastic injection molding product that adopts YOLOv5 algorithm in the deep learning neural network, including application of PAN and adaptive anchor boxes, which boosts the performance of small and medium-scale detectors and improves the performance and accuracy of bounding box prediction for localization.

[0010] It is also an objective of the present invention to provide a defect detection method and system with a deep learning model for plastic injection molding product, which provides a defect detection method and system with target region localization for at least one defect and at least one type of defect on at least one product of the same kind.

[0011] It is also a further objective of the present invention to provide a defect detection method and system with a deep learning model for plastic injection molding product, which provides a focus layer at the beginning and employing CSP to the feature map of the deep learning model for reducing network parameters and boosting the speed of the algorithm.

[0012] Accordingly, these objectives may be achieved by following the teachings of the present invention. The present invention relates to a defect detection method with machine learning for plastic injection molding product, comprising the steps of: collecting image dataset from product samples with data augmentation by adjusting environment and camera settings, and by a computational algorithm; training a model using the collected image dataset; evaluating the model; and detecting defects by using the evaluated model.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The features of the invention will be more readily understood and appreciated from the following detailed description when read in conjunction with the accompanying drawings of the preferred embodiment of the present invention, in which:

[0014] FIG.1 illustrates a method diagram of the defect detection with machine learning for plastic injection molding product in the present invention;

[0015] FIG.2 illustrates a model structure of YOLOv5 that is adopted to the deep learning model network in the present invention;

[0016] FIG.3 illustrates an example of existing model structure with the use of transfer learning;

[0017] FIG.4 illustrates an example of existing model structure with U-Net style;

[0018] FIG.5 illustrates an example of YOLOv3 structure;

[0019] FIG.6 illustrates an example comparing the performance of YOLOv3 on

Common Object in Context (COCO) dataset with other algorithms;

[0020] FIG.7 illustrates a single focus layer in YOLOv5; and

[0021] FIGS.8 illustrate an example of PANet architecture including FPN backbone, bottom-up path augmentation, and adaptive feature pooling.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0022] For the purposes of promoting and understanding of the principles of the invention, reference will now be made to the embodiments illustrated in the drawings and described in the following written specification. It is understood that the present invention includes any alterations and modifications to the illustrated embodiments and includes further applications of the principles of the invention as would normally occur to one skilled in the art to which the invention pertains.

[0023] The present invention teaches a defect detection method with machine learning for plastic injection molding product, comprising the steps of: collecting image dataset **100** from product samples with data augmentation by adjusting environment and camera settings **100a**, and by a computational algorithm **100c**; training a model **102** using the collected image dataset; evaluating the model **104**; and detecting defects **106** by using the evaluated model.

[0024] In a preferred embodiment of the present invention, the training of model **102** comprises the steps of: inputting the image dataset to a deep learning neural network for model training **102**; applying transfer learning **102b** to the deep learning neural network using the pre-trained dataset; applying model structure of YOLOv5 **102a** to the deep learning neural network; passing the inputted image to a single focus layer **102c** in the beginning layers of the deep learning neural network; wherein the single focus layer **102c** comprises of two convolution layers and a bottleneck layer; slicing each depth channel of the inputted image into four slices; concatenating the image in-depth slices; and applying another convolution layer with batch normalization and ReLU activation function.

[0025] In a preferred embodiment of the present invention, the training of model **102** further comprises the step of: duplicating feature map into two layers and merging the two layers together, before and after entering a dense block using cross-stage partial (CSP) networks **102e** when concatenating the in-depth slices of image.

[0026] In a preferred embodiment of the present invention, the training of model **102** further comprises the step of: adopting a path aggregation network (PAN) **102d** by the concatenation of semantical features and fine-grained features of the image.

[0027] In a preferred embodiment of the present invention, the training of model **102** further comprises the steps of: training the model **102** within a predefined number of consecutive epochs; and measuring metrices to monitor the performance of the trained model at each epoch.

[0028] In a preferred embodiment of the present invention, wherein the evaluating of the model **104** comprises the steps of: checking the presence of bias in the image dataset, presence of inaccurate annotation, reviewing and adjusting hyperparameters of the model; comparing the result after checking with a threshold to determine the accuracy of the model; routing the model for detecting defects **106** when the accuracy exceeds the threshold; repeating the collecting of data image dataset **100**, data annotation **100b** and adjusting of hyperparameters of the model when there is presence of bias, inaccurate and inappropriate hyperparameters; and applying early stopping to the model training **102** if a pre-defined epoch is reached or no improvement in the measured metrices is observed.

[0029] In a preferred embodiment of the present invention, the training of model **102** further comprises the steps of: training adaptive anchor boxes **106a**; and predicting a bounding box of products by using the trained adaptive anchor.

[0030] In a preferred embodiment of the present invention, the detecting defect **106** of the evaluated model comprises the step of; detecting at least one defect and at least one type of defect on at least one product of the same kind The detecting defect **106** of the evaluated model, further comprises the step of: labelling and localizing the defective region of the products based on the predicted bounding boxes.

[0031] The present invention also teaches a defect detection system for plastic injection molding product, comprising of: a data acquisition module to collect image dataset **100**; a model training **102** comprises a deep learning neural network that provides instructions on: applying transfer learning **102b** to the deep learning neural network; applying model structure of YOLOv5 **102a** to the deep learning neural network; passing the image to a single focus layer **102c** in the beginning layers of the deep learning neural network; slicing each depth channel of the image; concatenating the image slice in depth using cross-stage partial (CSP) networks **102e**; applying a convolution layer with batch normalization and a ReLU activation function; adopting path aggregation network (PAN) **102d**; training adaptive anchor boxes **106a**; an evaluation module for testing the accuracy of the model **104a**; and a defect detection module to detect at least one defect and at least one type of defect on at least one product of the same kind.

[0032] In a preferred embodiment of the present invention, the data acquisition module comprises a data augmentation by adjusting environment and camera settings **100a**, and by a computational algorithm **100c**.

EXAMPLE

[0033] A diagram of the defect detection method with machine learning for plastic injection molding product in the present invention is illustrated in **FIG.1**. The method comprises the steps of: collecting image dataset **100** from product samples; training a model **102** using the collected image dataset; evaluating the model **104**; and detecting defects **106** by using the evaluated model. The said training of model **102** further comprises the steps of: inputting the image dataset to a deep learning neural network for model training **102**; applying transfer learning **102b** to the deep learning neural network using the pre-trained dataset; applying model structure of YOLOv5 **102a** to the deep learning neural network; passing the image to a single focus layer **102c** in the beginning layers of the deep learning neural network; slicing each depth channel of the image into slices; concatenating the image in-depth slices using CSP networks **102e**; adopting PAN **102d**; and training the adaptive anchor boxes **106a**. The model will then use the trained adaptive anchor boxes **106a** to make prediction.

[0034] As further illustrated in **FIG.1**, the model is evaluated after training to test the accuracy of the model **104a**, wherein the image dataset and/or hyperparameters of the model will be reviewed if the accuracy performance is unacceptable. Accordingly, the defective region of the products is labelled and localized from the prediction of the trained adaptive anchor boxes **106a** during training. The model structure of YOLOv5 **102a** adopted into the deep learning neural network is illustrated in **FIG.2**.

[0035] **FIG.3** shows the existing model structure with transfer learning **102b** wherein only comprises feature extraction and image classification, **FIG.4** shows another example of the existing model structure with U-Net style for predicting the class probability on each pixel. Another example of the YOLOv3 structure is shown in **FIG.5**. The performance of the said YOLOv3 on the Common Object in Context (COCO) dataset is compared with other algorithms, as shown in **FIG.6**.

[0036] The single focus layer **102c** in the beginning layers of the deep learning neural network in the present invention is combined with two convolution layers, followed by a bottleneck layer. The method further comprises the step of: slicing each depth channel of the image into four slices with a step size of two as shown in **FIG.7**. The size of the slices is reduced by half from the input image size. Then, the 12 slices are further concatenated in-depth, followed by a convolution layer with batch normalization and a ReLU activation function. The focus layer **102c** acts like down-sampling in the beginning while information is reserved. Therefore, the first few layers from the old YOLO series are replaced by the focus layer **102c** and hence the learning parameters are reduced in the present invention. Therefore, the speed of the algorithm can be boosted.

[0037] The use of PAN **102d** in the present invention is an improved technique over the feature pyramid network (FPN) used in YOLOv3. FPN is applied in YOLOv3 for detecting small objects in the large-scale detector. As shown in **FIG.8**, the complexity of semantical features increases along with the high-level layers from the bottom layer to the top layer. As a trade-off, the fine-grained features (spatial information) are further decreased in high-level layers due to multiple processes of down-sampling. A top-down path of semantical features is then applied in FPN to preserve the fine-grained features by concatenating both semantical features and fine-grained features. This allows the detection of small objects by large-scale detectors. However, fine-grained features in low-level layers are not concatenated with the semantical features in high-level layers. Therefore, poor fine-grained features are used for detection in small-scale detectors and hence the accuracy in small-scale object detection is limited. To improve the performance of small-scale detectors of high-level layers in YOLOv5 **102a**, a bottom-up path is applied in PAN **102d** by the concatenation of semantical features and fine-grained features, as further illustrated in **FIG.8**. The new path from low-level layers to high-level layers in PAN **102d** involves less than ten layers, whereas the old path involves many more layers in-between the backbone layers. With the PAN **102d** method, a better combination between semantical features and fine-grained features is achieved in small-to-medium-scale detectors. Therefore, the accuracy of detection of small-to-medium-scale objects is improved.

[0038] The backbone of YOLOv5 **102a** in the present invention employs the strategy from CSP networks **102e** in duplicating the feature map into two layers and then merging them together at different stages. The use of CSP networks **102e** can reduce the training parameters significantly by integrating gradient changes in feature map. Huge amounts of gradient information are reused for updates when concatenation is involved, while the repeated information becomes redundant. The CSP networks **102e** can solve this issue effectively by integrating the gradient changes from dense layers separately. Therefore, the network parameters are reduced, while the speed of this invention can be further boosted.

[0039] In the YOLO series, the anchor box is a useful tool for predicting the bounding box of detected objects. In YOLOv3, initial anchor boxes are obtained by running the k-means clustering on the dataset, while bounding boxes are predicted from those pre-defined anchor boxes with shifting and scaling. However, the pre-defined anchor boxes may have a high error rate if the aspect ratio of the object in the dataset is highly variable, while the prediction of the bounding box afterwards will be affected. To improve the performance of anchor boxes, adaptive anchor boxes **106a** are applied in YOLOv5 **102a** in the present invention. During model training **102**, anchor boxes are re-defined by k-means clustering once their best recall is lower than some pre-defined threshold. This enhancement improves the anchor boxes' suitability over time. As a result, the accuracy of the bounding box prediction in YOLOv5 **102a** can be improved.

[0040] The YOLOv5 **102a** is used to detect objects, and the defective region can be labelled with a bounding box to show it clearly. Moreover, more than one sample or a sample with more than one defect can also be detected in the present invention with localization. This can boost the detection efficiency as a batch of products can be detected together. This can also alleviate the issue of limited computational power. If the computational power is not strong enough, the frames per second (FPS) of the detection may be reduced. As a result, the product's detection rate will be reduced directly. By using the method of the present invention with multi-objects detection, products can be grouped in a batch to be detected, and the detection rate of the products can be increased under the same FPS condition.

[0041] In addition, the training of model **102** in the present invention is terminated once the pre-defined epoch is reached or no improvement in the measured metrices is observed after the pre-defined number of consecutive epochs. The addition of early stopping to the model training **102** could prevent the model from over-fitting in order to improve prediction performance.

[0042] In the present invention, evaluation of the model **104** is carried out after the training of the model **102**. The dataset obtained from the new environment setting, together with new types of defective samples that were not used for model training **102**, is used for evaluation. If the evaluation of model **104** accuracy exceeds a certain threshold, such as 0.95, it is routed to the detection of defects. The evaluation module includes but not limited to checking any bias in the dataset, checking any inaccurate annotation, reviewing and adjusting hyperparameters of the model, if required. The image dataset will be recollected if bias is found, such as insufficient samples in a particular class or environment. If inaccurate annotation is discovered in the data acquisition module, such as a mislabelled class or an incorrect bounding box, data annotation **100b** must be repeated. The performance of the method and system is guaranteed by the evaluation module with the accurate criteria, which will proceed to the defect detection module.

[0043] The present invention explained above is not limited to the aforementioned embodiment and drawings, and it will be obvious to those having an ordinary skill in the art of the present invention that various replacements, deformations, and changes may be made without departing from the scope of the invention.

CLAIMS

WHAT IS CLAIMED:

1. A defect detection method with machine learning for plastic injection molding product, comprising the steps of:

collecting image dataset (100) from product samples with data augmentation by adjusting environment and camera settings (100a), and by a computational algorithm (100c);

training a model (102) using the collected image dataset;

evaluating the model (104); and

detecting defects (106) by using the evaluated model.

2. The defect detection method with machine learning for plastic injection molding product, according to claim 1, wherein the training of model (102) comprises the steps of:

inputting the image dataset to a deep learning neural network for model training (102);

applying transfer learning (102b) to the deep learning neural network using the pre-trained dataset;

applying model structure of YOLOv5 (102a) to the deep learning neural network;

passing the inputted image to a single focus layer (102c) in the beginning layers of the deep learning neural network;

wherein the single focus layer (102c) comprises of two convolution layers and a bottleneck layer;

slicing each depth channel of the inputted image into four slices;

concatenating the image in-depth slices; and

applying another convolution layer with batch normalization and ReLU activation function.

3. The defect detection method with machine learning for plastic injection molding product, according to claim 2, wherein the training of model (102) further comprises the steps of:

duplicating feature map into two layers and merging the two layers together, before and after entering a dense block using Cross-Stage Partial (CSP) networks (102e) when concatenating the in-depth slices of image.

4. The defect detection method with machine learning for plastic injection molding product, according to claim 3, wherein the training of model (102) further comprises the step of:

   adopting Path Aggregation Network (PAN) (102d) by the concatenation of semantical features and fine-grained features of the image.

5. The defect detection method with machine learning for plastic injection molding product, according to claim 4, wherein the training of model (102) further comprises the steps of:

   training the model (102) within a predefined number of consecutive epochs; and

   measuring metrices to monitor the performance of the trained model at each epoch.

6. The defect detection method with machine learning for plastic injection molding product, according to claim 5, wherein the evaluating of the model (104) comprises the steps of:

   checking the presence of bias in the image dataset, presence of inaccurate annotation, reviewing and adjusting hyperparameters of the model;

   comparing the result after checking with a threshold to determine the accuracy of the model;

   routing the model for detecting defects (106) when the accuracy exceeds the threshold;

   repeating the collecting of data image dataset (100), data annotation (100b) and adjusting of hyperparameters of the model when there is presence of bias, inaccurate and inappropriate hyperparameters; and

   applying early stopping to the model training (102) if a pre-defined epoch is reached or no improvement in the measured metrices is observed.

7. The defect detection method with machine learning for plastic injection molding

product, according to claim 5 or 6, wherein the training of model (102) further comprises the steps of:

training adaptive anchor boxes (106a); and

predicting a bounding box of products by using the trained adaptive anchor.

8. The defect detection method with machine learning for plastic injection molding product, according to claim 7, wherein the detecting defect (106) of the evaluated model, comprises the step of:

detecting at least one defect and at least one type of defect on at least one product of the same kind.

9. The defect detection method with machine learning for plastic injection molding product, according to claim 8, wherein the detecting defect (106) of the evaluated model, further comprises the step of:

labelling and localizing the defective region of the products based on the predicted bounding boxes.

10. A defect detection system for plastic injection molding product comprises of:

a data acquisition module to collect image dataset (100);

a model training (102) comprises a deep learning neural network that provides instructions on:

applying transfer learning (102b) to the deep learning neural network;

applying model structure of YOLOv5 (102a) to the deep learning neural network;

passing the image to a single focus layer (102c) in the beginning layers of the deep learning neural network;

slicing each depth channel of the image;

concatenating the image slice in depth using Cross-Stage Partial (CSP) networks (102e);

applying a convolution layer with batch normalization and a ReLU activation function;

adopting path aggregation network (PAN) (102d);

training adaptive anchor boxes (106a);

an evaluation module for testing the accuracy of the model (104a); and

a defect detection module to detect at least one defect and at least one type of defect on at least one product of the same kind.

11. The defect detection system for plastic injection molding product, according to claim 10, wherein the data acquisition module comprises a data augmentation by adjusting environment and camera settings (100a), and by a computational algorithm (100c).

FIG. 1

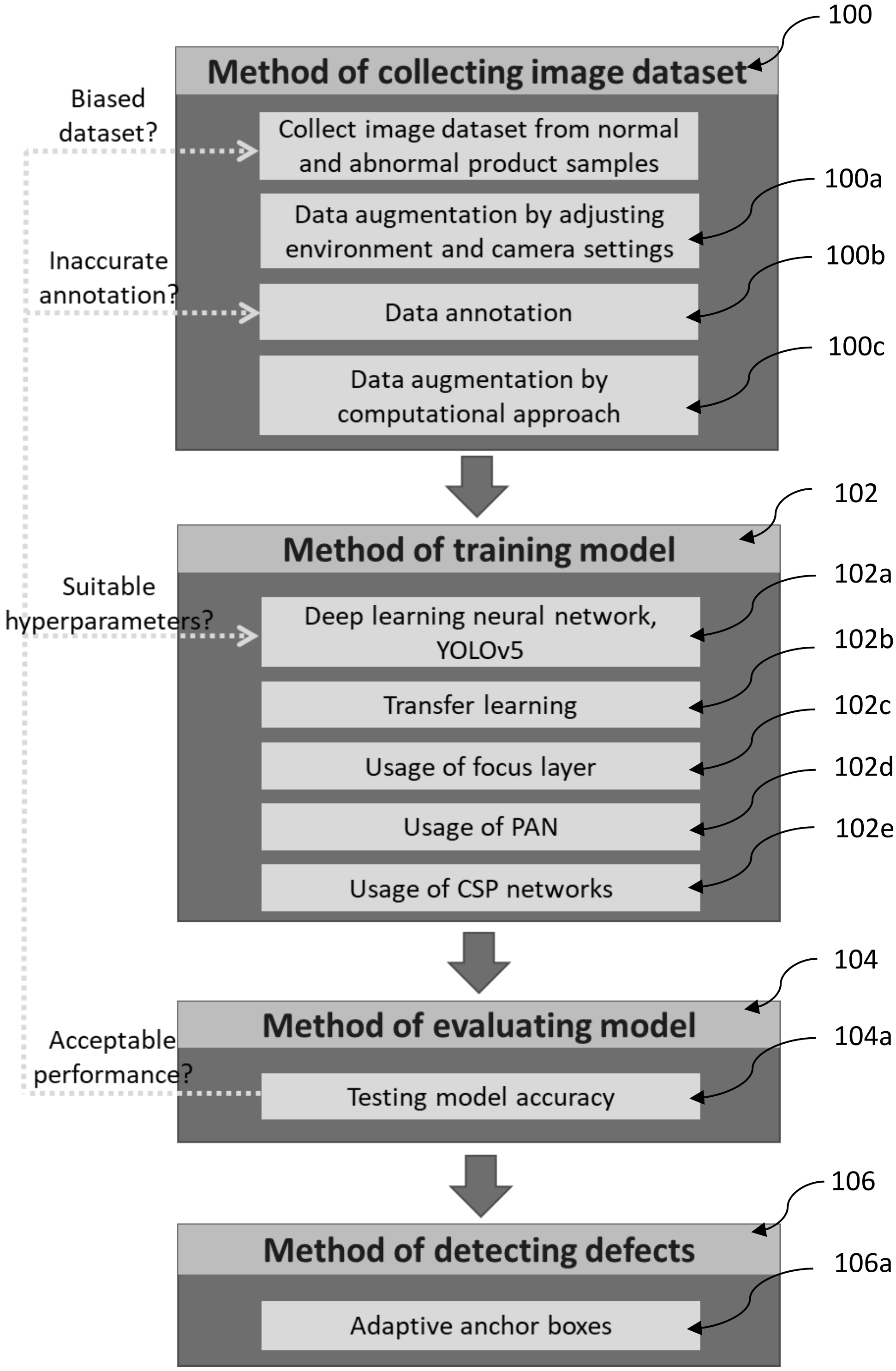
100
Method of collecting image dataset
Biased dataset?
Collect image dataset from normal and abnormal product samples
100a
Data augmentation by adjusting environment and camera settings
100b
Inaccurate annotation?
Data annotation
100c
Data augmentation by computational approach
102
Method of training model
102a
Suitable hyperparameters?
Deep learning neural network, YOLOv5
102b
Transfer learning
102c
Usage of focus layer
102d
Usage of PAN
102e
Usage of CSP networks
104
Method of evaluating model
104a
Acceptable performance?
Testing model accuracy
106
Method of detecting defects
106a
Adaptive anchor boxes

FIG. 1

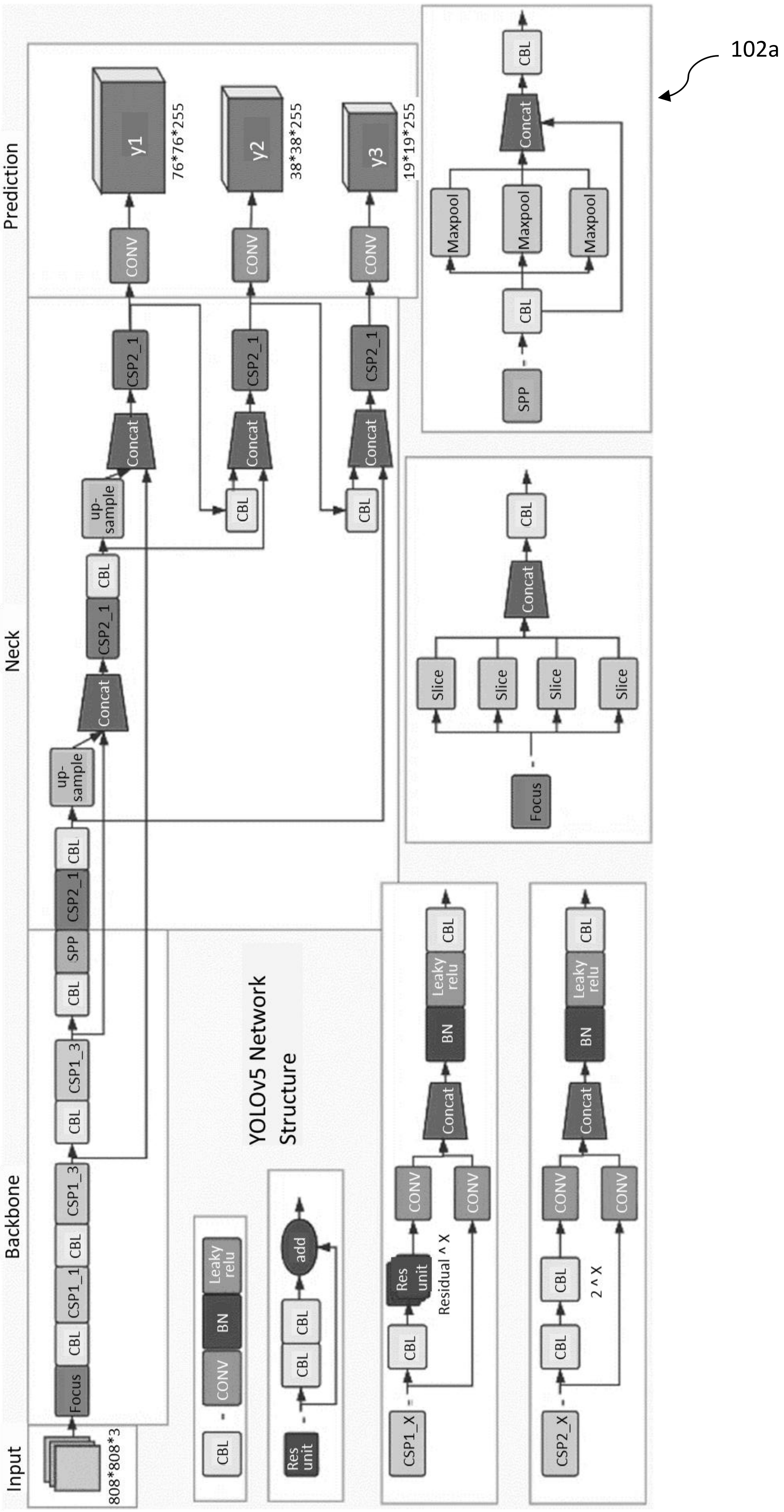
Input
Backbone
Neck
Prediction
808*808*3
Focus
CBL
CSP1_1
CSP1_3
SPP
CSP2_1
up-sample
Concat
CONV
y1
y2
y3
76*76*255
38*38*255
19*19*255
YOLOv5 Network Structure
BN
Leaky relu
Res unit
add
CSP1_X
Residual ^ X
CSP2_X
2 ^ X
Slice
Maxpool
102a

FIG.2

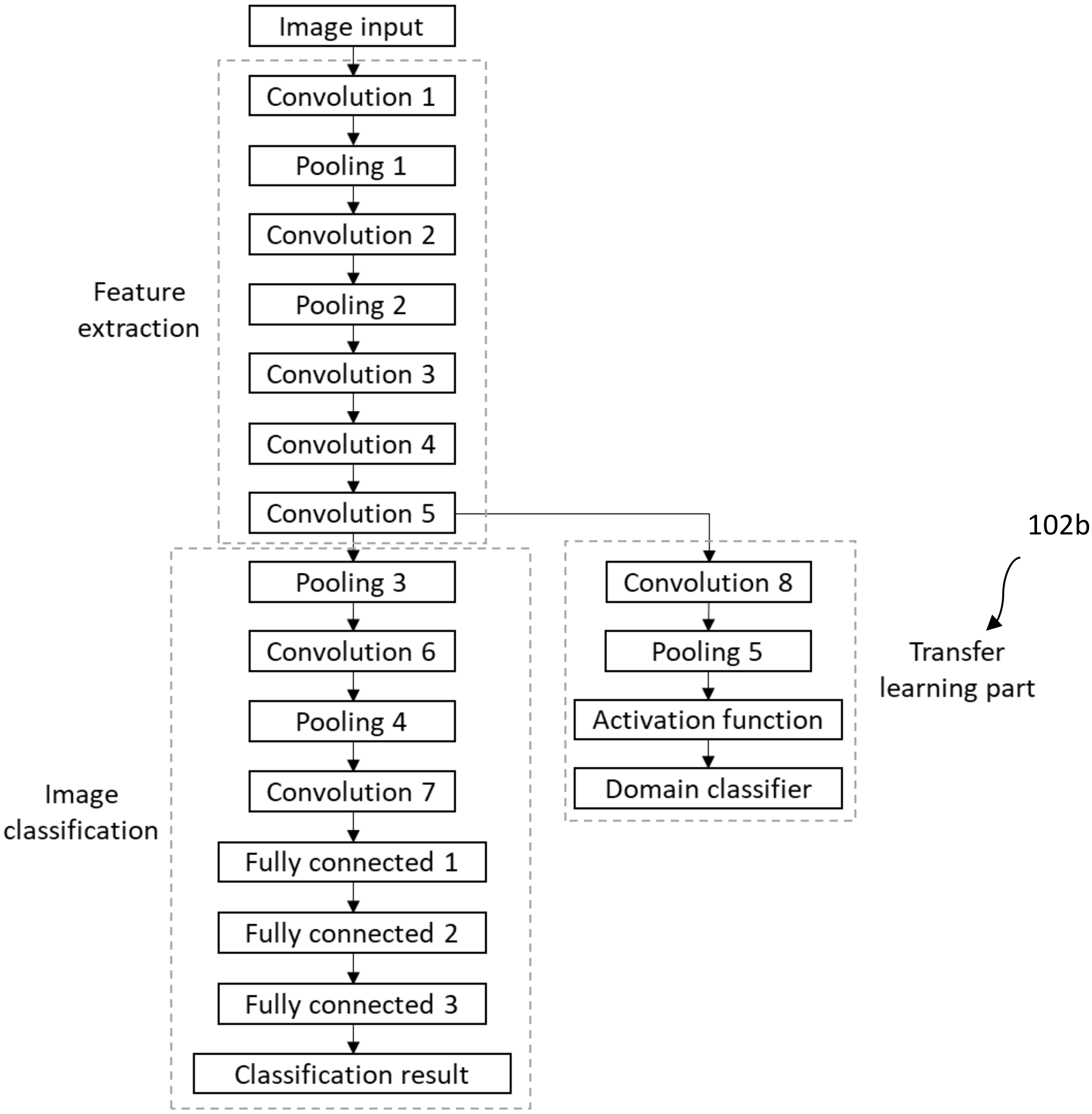
Image input
Convolution 1
Pooling 1
Convolution 2
Pooling 2
Convolution 3
Convolution 4
Convolution 5
Feature extraction
Pooling 3
Convolution 6
Pooling 4
Convolution 7
Fully connected 1
Fully connected 2
Fully connected 3
Classification result
Image classification
Convolution 8
Pooling 5
Activation function
Domain classifier
102b
Transfer learning part

FIG. 3

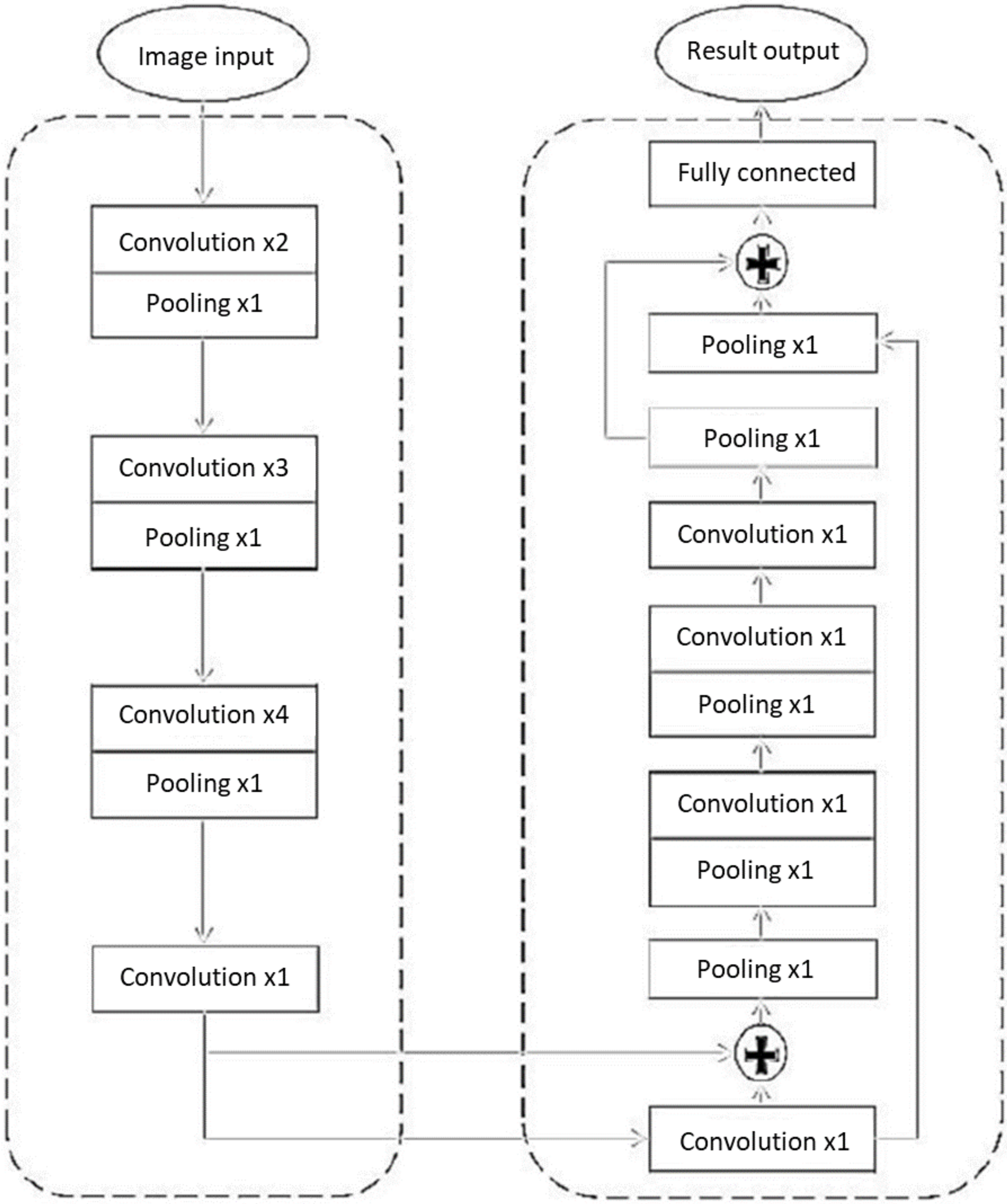
Image input
Convolution x2
Pooling x1
Convolution x3
Pooling x1
Convolution x4
Pooling x1
Convolution x1
Result output
Fully connected
Pooling x1
Pooling x1
Convolution x1
Convolution x1
Pooling x1
Convolution x1
Pooling x1
Pooling x1
Convolution x1

FIG. 4

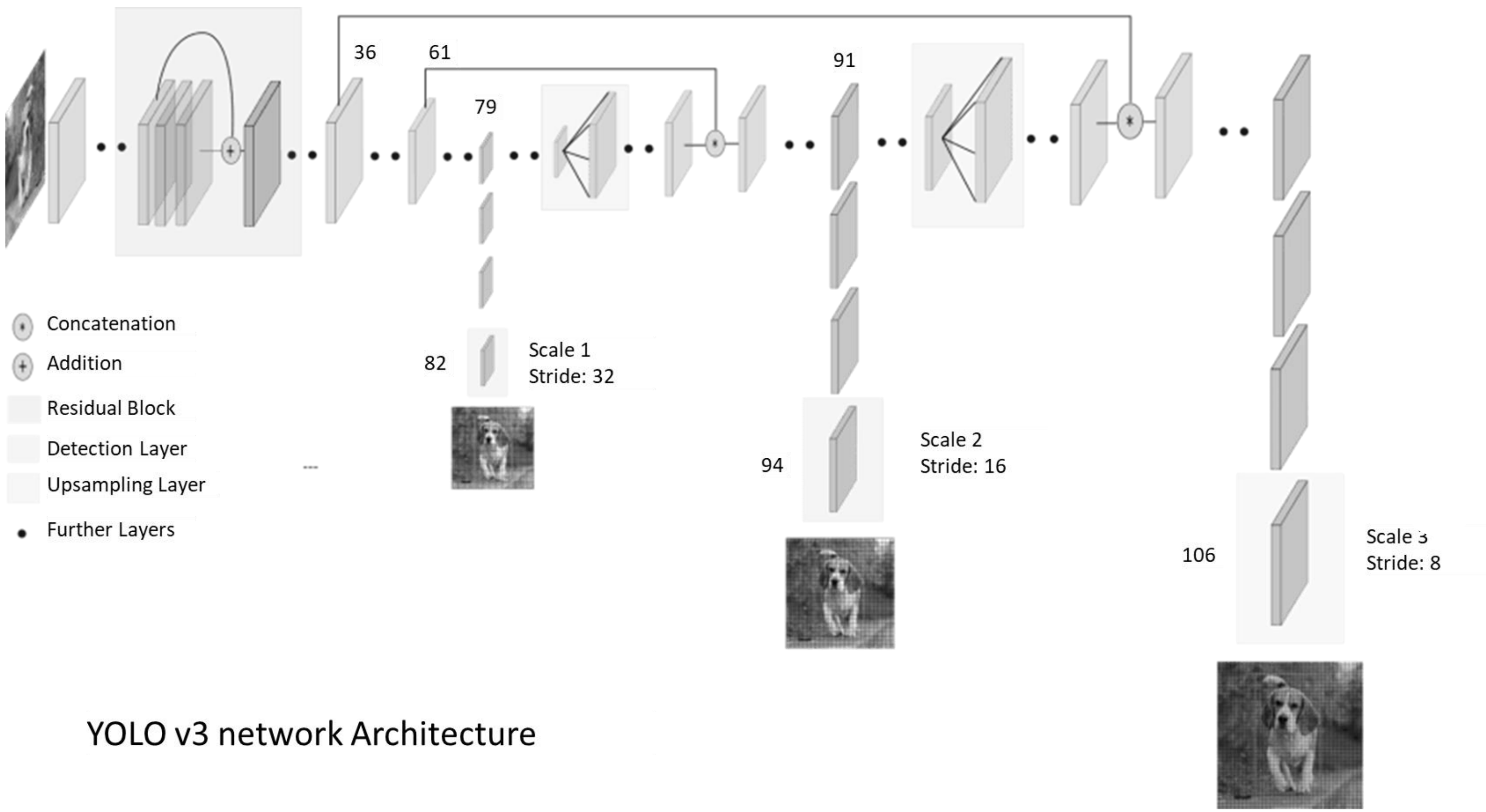
36
61
79
91
Concatenation
Addition
Residual Block
Detection Layer
Upsampling Layer
Further Layers
82
Scale 1
Stride: 32
94
Scale 2
Stride: 16
106
Scale 3
Stride: 8
YOLO v3 network Architecture

FIG.5

	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI	Inception-ResNet-v2	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	**52.1**
One-stage methods							
YOLOv2	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD513	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet	ResNeXt-101-FPN	**40.8**	**61.1**	**44.1**	**24.1**	**44.2**	51.2
YOLOv3 608x608	DarkNet-53	33.0	57.9	34.4	18.3	35.4	41.9

FIG. 6

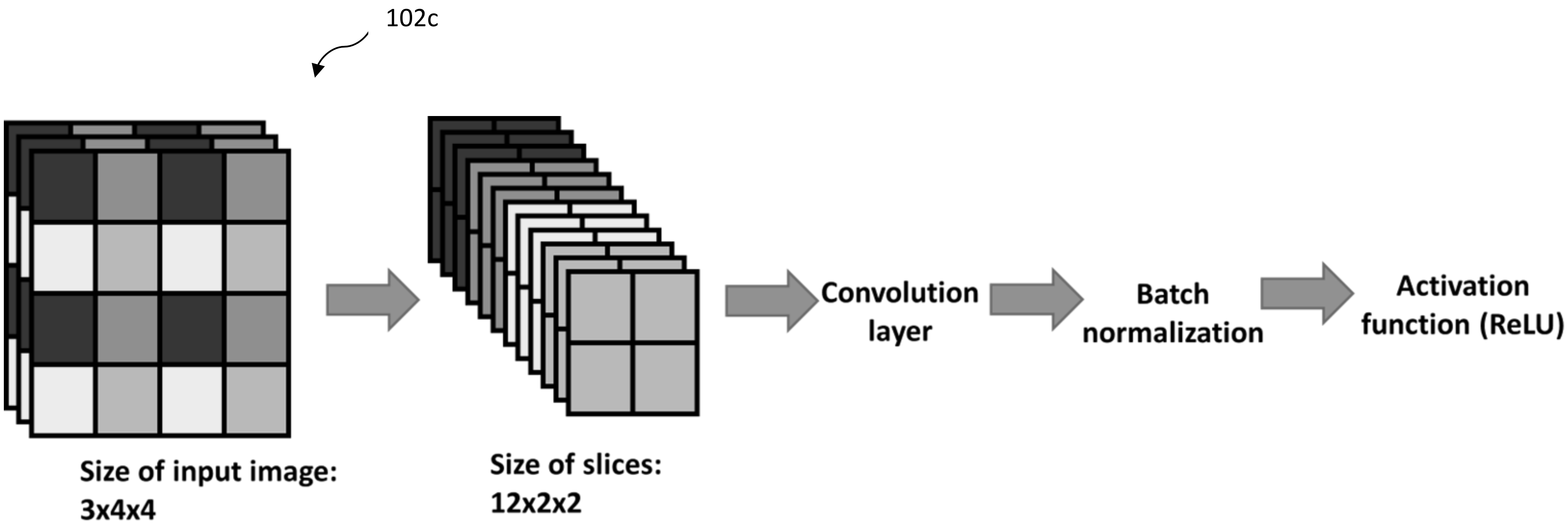
102c
Size of input image:
3x4x4
Size of slices:
12x2x2
Convolution layer
Batch normalization
Activation function (ReLU)

FIG.7

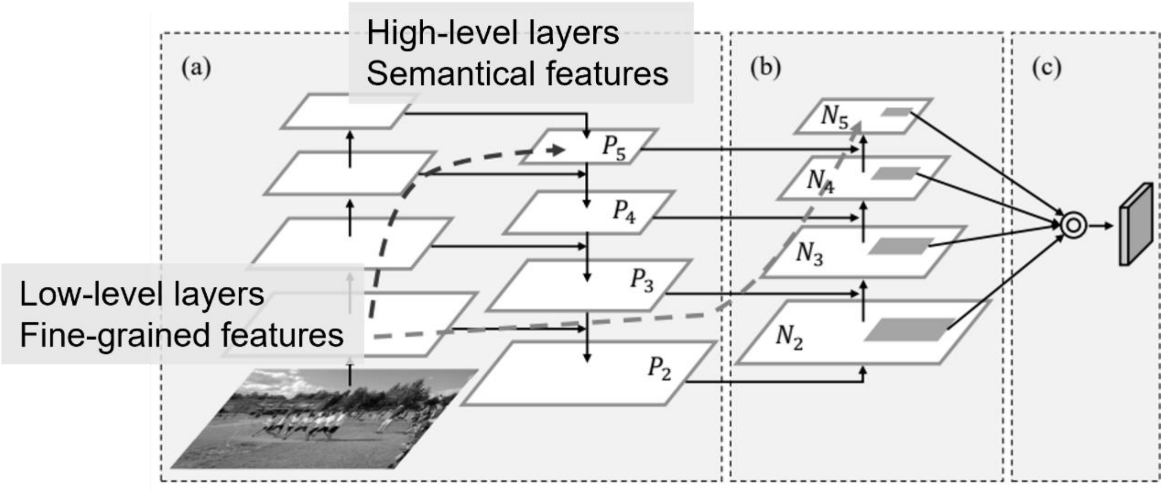
High-level layers
Semantical features
(a)
(b)
(c)
P_5
P_4
P_3
P_2
N_5
N_4
N_3
N_2
Low-level layers
Fine-grained features

FIG.8