applied sciences

MDPI

Article

Classification of the Trap-Neuter-Return Surgery Images of Stray Animals Using Yolo-Based Deep Learning Integrated with a Majority Voting System

Yi-Cheng Huang [1,*], Ting-Hsueh Chuang [2] and Yeong-Lin Lai [2]

[1] Department of Mechanical Engineering, National Chung Hsing University, Changhua 50074, Taiwan
[2] Department of Mechatronics Engineering, National Changhua University of Education, Changhua 50007, Taiwan; d0651003@mail.ncue.edu.tw (T.-H.C.); yllai@cc.ncue.edu.tw (Y.-L.L.)
* Correspondence: ychuang66@dragon.nchu.edu.tw

Abstract: Trap-neuter-return (TNR) has become an effective solution to reduce the prevalence of stray animals. Due to the non-culling policy for stray cats and dogs since 2017, there is a great demand for the sterilization of cats and dogs in Taiwan. In 2020, Heart of Taiwan Animal Care (HOTAC) had more than 32,000 cases of neutered cats and dogs. HOTAC needs to take pictures to record the ears and excised organs of each neutered cat or dog from different veterinary hospitals. The correctness of the archived medical photos and the different shooting and imaging angles from different veterinary hospitals must be carefully reviewed by human professionals. To reduce the cost of manual review, Yolo's ensemble learning based on deep learning and a majority voting system can effectively identify TNR surgical images, save 80% of the labor force, and its average accuracy (mAP) exceeds 90%. The best feature extraction based on the Yolo model is Yolov4, whose mAP reaches 91.99%, and the result is integrated into the voting classification. Experimental results show that compared with the previous manual work, it can decrease the workload by more than 80%.

Keywords: trap-neuter-return (TNR); object detection; multi-classifier majority voting; TNR surgery images

Citation: Huang, Y.-C.; Chuang, T.-H.; Lai, Y.-L. Classification of the Trap-Neuter-Return Surgery Images of Stray Animals Using Yolo-Based Deep Learning Integrated with a Majority Voting System. *Appl. Sci.* **2021**, *11*, 8578. https://doi.org/10.3390/app11188578

Academic Editor: Joonki Paik

Received: 6 August 2021
Accepted: 13 September 2021
Published: 15 September 2021

1. Introduction

1.1. Backgruond

The issue of stray dogs and cats has become a global concern because of animal welfare and public hygiene; moreover, stray dogs and cats have an impact on native wildlife through predation, competition, and disease transmission [1,2]. To reduce the number of stray animals, people first impound them into shelters; however, more than half of impounded animals are euthanized due to shelter crowding, infectious disease, and feral behavior [3]. Stray animal euthanasia has been used for many years, but it is not able to decrease the number of stray animals effectively. Besides, there are factors showing that people performing the euthanasia often face serious mental problems such as depression, insomnia, and even suicide [4–7]. Therefore, the trap-neuter-return (TNR) strategy has grown in recent decades [8–13]. TNR, a non-lethal alternative to control the number of stray cats, can be traced back to the 1950s in England. Since then, it has spread to many countries in the world [14]. The Heart of Taiwan Animal Care Association (HTACA), one of Taiwan's animal protection groups, is a non-profit, non-governmental civil organization that puts efforts in animal protection, animal life education courses, rural sterilization operations, stray dog and cat rescue and sterilization programs, animal shelters, and university animal protection activities. Due to the culling-free policy for stray cats and dogs since 2017, Taiwan has a large demand for cat and dog sterilization. In 2020, the HTACA alone had around 32,000 neutered cases of cats and dogs. Each case has one or two photos recorded, which will increase year by year. Every dog or cat neutered from veterinary hospitals must be photographed, including its ears and excised organs as the corresponding voucher for

each dog or cat surgery. The problem is that the up to 60,000 archived medical treatment photos from different veterinary hospitals and different shooting imaging angles must be accurate. Since there is no special camera equipment or photography specifications for each hospital, images have to be checked manually and reviewed carefully by a professional human's eye.

Neural networks are models with deep structures (multiple layers of neuron structures) that have been developed for a long time and became popular in the 1980s and 1990s [15]. However, due to the limitations of computing power and datasets, they encountered a bottleneck in accuracy and speed around the year 2000, coupled with the rise of the support vector machine (SVM). Since 2005, the emergence of large, annotated datasets, such as pattern analysis, statistical modeling, and computational learning (PASCAL) [16], ImageNet Large-Scale Visual Recognition Competition, (ILSVR) [17], and the development of high-performance computing technologies like processor clusters and GPUs, deep models have shown efficient capabilities in many published models, such as ResNet [18], VGG16 [19]. Girshick et al. [20] proposed region-based convolutional neural networks (R-CNNs), which are defined by a two-stage architecture by integrating a region-wise and object detector for object detection.

The region-wise feature extraction ideas of SPP-Net [21] and Fast R-CNN [22] greatly accelerate the speed of the entire detector. Ren et al. [23] introduce a region proposal network (RPN) [22] that shares the full-image convolution feature with the detection network, thereby realizing almost cost-free region proposals. Cascade R-CNN [24] is a multi-stage object detection architecture that improves accuracy by a sequence of detectors trained with increasing intersection over union (IoU) thresholds. In the network architecture of object detection, one-stage object detection (a neural network that can detect the position of objects and identify objects at the same time) is very popular due to its computational efficiency. Yolo [25] is implemented through an effective backbone network and supports real-time object detection. Yolov3 uses Darknet-53 [26] as the backbone network. Yolov4 [27] uses CSPDarknet-53 [28] as the backbone network. Since the feature pyramid network (FPN) [29] performs well in solving multi-scale problems, Yolov3 and Yolov4 improve network performance by integrating FPN and the Residue Network (ResNet). The Single Shot MultiBox Detector (SSD) [30] uses multiple feature maps with multiple resolutions to identify objects of different proportions and detect objects in a method that is similar to RPN. Most two-stage architectures are higher than the faster first-stage detectors. However, by solving the problem of foreground and background imbalances in dense object detection, RetinaNet [31] achieved better results with more accuracy than most two-stage object detectors. This is because this algorithm solves the foreground-background imbalance in dense object detection. The masked region convolutional neural network (Mask-RCNN) proposed by Kaiming et al. [32] has been able to integrate object detection and instance segmentation into a single framework. However, in this study, we need to locate the surgical location of the image. The classification neural network recognition area cannot be located. The object detection network's predictive output is the bounding box, so the object area can be located clearly. The time for the person to work the image label is within 1 min (each picture of our research case requires four label areas). Semantic recognition technology can accurately confirm the contours of surgical organs but requires more image label time (in this case, it may take 3 to 5 min). Labeling the contours of surgical organs is a tedious task (as shown in Table 1). Thus, we chose the object detection network. In the object detection neural network, we use the Yolo series of neural networks. YOLOv4 obtained an AP value of 43.5% on the MS COCO dataset and achieved a real-time speed up to 65 FPS on the Tesla V100, gaining the title of the fastest and the most accurate detector among the Yolo series. The Yolo series neural network has many network resources. For edge computing applications and neural network modifications, the network resources are relatively rich. So, this research uses Yolov3 tiny, Yolov3, and Yolov4 to fetch the key features of image samples.

Table 1. Neural network selection evaluation form.

	Classification Neural Network	Object Detection Network	Semantic Recognition Neural Network
Can it locate the object area?	No	Yes	Yes
Image label time	Short	within 1 min	3 to 5 min

1.2. Related Work

In recent years, using remote still and video surveillance in the research of wildlife and management has grown rapidly [33–35]. Surveillance has many purposes, from the identification of pests or species with problem behaviors to the assessment of species diversity and distribution, and thus the importance of conservation. However, these usually share a common need, which is to identify a specific target species. In one study, they used image recognition as an estimation tool, preventing Mycobacterium bovis transmission between badgers and cattle. With growing interest in remote surveillance, capturing large amounts of image data becomes a challenge in identification. Before, screening pictures was required to be quality checked manually, which is expensive and time consuming [36,37]. Thus, people are deeply interested in the emergence of automated methods [35]. Machine learning has an increasing trend in automatic animal identification, and it has also been applied in biological and fishery monitoring. These technologies have made great progress, and can shoot high-resolution images in challenging environments, and in the end can effectively manage natural resources [38]. However, the method of identifying animals depends on the situation. For instance, it has been successful when using colors to distinguish animals from the background in automatic detection and tracking of elephants. Furthermore, it also works on other species [35]. If there are a lot of visual data including target and non-target species that require manual review, machine learning that can automatically identify wildlife needs to be used. When it comes to classification, image features are required. Generally, the histogram of oriented gradient [39], one of hand-crafted image feature methods, has been widely applied. Scale-invariant feature transform (SIFT) is also another hand-crafted image feature method [40]. However, convolutional neural networks (CNN) [41,42] perform better than all hand-crafted feature methods on large datasets [43]. For example, the ImageNet dataset includes 1.2 million images. In recent years, it has been applied to the automatic classification of wildlife, but its performance still has its limitations [37]. In Reference [44], the accuracy of deep learning classification photos was evaluated with the processing of a few collections of picture materials that are difficult to model. Their trained model was applied on a project from an online citizen science platform that provides researchers with access to millions of volunteers via www.zooniverse.org, where public volunteers can help ecologists classify large amounts of image data. Therefore, CNN was applied to differentiate among different species, humans, or vehicles and empty pictures (no animals, vehicles, or humans). Combining a trained model that has been classified by citizen scientists, manual work has been reduced by 43% while maintaining entire accuracy for a live experiment working on Zooniverse. It means that the trained neural network can correctly replace 43% of the human workload, and the result is the same as that done by humans. Besides, deep learning is used to automatically identify and separate species' specific activities from still images and video data using CNN. A total of 8368 wild and domestic animal images were used to develop a method to separate badgers from other species (two classifications), and then distinguish each from six species (multi-classification). This means making two classifications first to determine whether it is a target image, and if it is judged to be a target, then performing six classifications. In Reference [37], two deep learning frameworks are used to do automatic image recognition and get a high accuracy, 98.05%, for binary classification, and 90.32% for multi-classification. In 2017, CNN was used to train for recognizing and counting the behavior of 48 species in 3.2 million images [45]. The accuracy rate of neural network identification is greater than 93.8%, and the number of identifications is expected to increase rapidly in the years to

come. More importantly, if this system only classifies reliable images, it can automatically recognize 99.3% of the data, which saves more than 8.4 years of time compared with manual work while the classification rate of volunteers is still at 96.6% accuracy.

The goal of this paper is to deploy a deep learning method to save the previous time from doing this task manually. No research has been found yet on how to detect whether the animals have been neutered from the photos. A special challenge here is that it is not easy to obtain images of neutered animals, and there are certain differences among the images. Here, our goal is to develop a powerful framework to classify the images of neutered animals by using the method of "You Only Look Once" (Yolo) and ensemble learning. Deep learning is applicable to significant computational resources, large amounts of labeled data, and modern neural network architecture. We found that the application of deep learning has not been used in the neutered operation of TNR, but our experiments prove that the integrated algorithms of deep learning and majority voting methods have a good recognition rate for neutered operation images. Here, we combined thousands of photos with labeled data from the neutering stage in TNR, a modern GPU server, using Yolo-based deep learning and an ensemble learning of a majority voting system to test how to use deep learning to automatically detect the photos in the neutering stage of TNR. We found that the system performed as well as the human volunteer team on most of the data and identified a few images that required manual evaluation. The goal of the overall system is to automatically classify the sterilization images of male dogs, female dogs, male cats, and female cats. The best feature extraction by the Yolo-based model is Yolov4, whose mAP reached 91.99%, and the majority voting system reached the accuracy of 85%, so it can save more than 86% of manual detection time. Experimental results show it can save more than 80% time in comparison with the previous manual work.

2. Methodology

2.1. Yolov3 Deep Learning Network

The feature extraction network of Yolov3 is Darknet-53, and its structure is similar to ResNet. The basic unit of Darknet-53 using 1×1 and 3×3 convolutional layers and the remaining module. Darknet-53 uses the concept of Shortcut in ResNet to combine early feature maps with up-sampling feature maps. It can combine the coarse-grained features of the early stage with the fine-grained features of the later stage, so that the entire feature extraction can capture more comprehensive features. Darknet-53 retains the leaked ReLU layer and Batch normalization layer. In addition, Darknet-53 also mentioned the concept of multi-scale feature layers in FPN, and selects the last three scale layers as the output, as shown in Figure 1. The loss function of the Yolov3 is composed of l_box, l_obj, and l_class terms. l_box is the loss brought by bounding boxes, and l_obj is the error caused by confidence. The last term l_class is the error caused by the category.

The loss function is defined as follows.

$$\text{Loss} = \sum_{i=0}^{s^2} \text{l}_{\text{box}} + \text{l}_{\text{obj}} + \text{l}_{\text{class}}$$

$$\text{, where } \text{l}_{\text{box}} = \sum_{i=0}^{s^2}\sum_{j=0}^{B}\left\{ l_{ij}^{obj} * \left[\text{l}_{\text{BCE}(x_i,\hat{x}_i)} + \text{l}_{\text{BCE}(y_i,\hat{y}_i)}\right]\right\} + \lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}\left\{ I_{ij}^{obj} * \left[(w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2\right]\right\}$$

$$\text{l}_{\text{obj}} = \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{obj}\left[\text{l}_{\text{BCE}(c_i,\hat{c}_i)}\right] + \lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{noobj}\left[\text{l}_{\text{BCE}(c_i,\hat{c}_i)}\right],$$

$$\text{l}_{\text{class}} = \sum_{i=0}^{s^2} I_i^{obj} \sum_{c \in classes} \text{lBCE}(p_i(c), \hat{p}_i(c)), \tag{1}$$

where

$$\text{l}_{\text{BCE}(a,\hat{a})} = -[a \log a + (1-a)\log(1-\hat{a})] \tag{2}$$

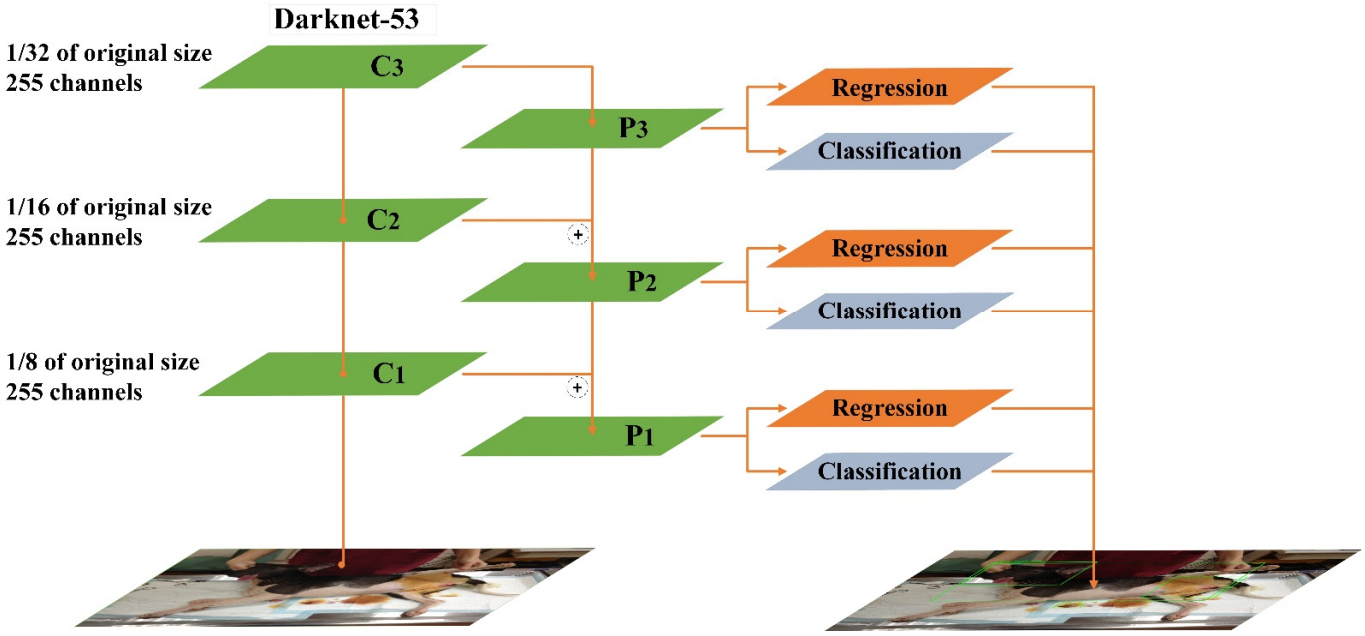


Figure 1. Structure of YOLOv3 [26] C1–C3 refer to the feature layers of the last three scale. P1–P3 (After superimposition of adjacent layers through upper sampling). ⊕ refers (The superimposition Operation).

The l_BCE function (2) is the cross entropy of loss function. Cross entropy is used to calculate the loss, and only applies to the grid with a target.

The 1_{ij}^{obj} points out that the target is detected by the jth bounding box of grid i. I_{ij}^{noobj} shows that there are no targets in this bounding box. B indicates that each grid predicts B bounding boxes, to raise the loss from bounding box coordinate predictions and reduce the loss for confidence predictions for boxes that do not contain objects. The parameters λcoord and λnoobj are brought in and set to 5. $\left(\hat{x}, \hat{y}, \hat{w}, \hat{h}, \hat{c}, \hat{p}\right)$ are respectively expressed as the center coordinates, width, height, confidence, and category probability of the predicted bounding box. Furthermore, those symbols without the cusp are the true labels.

2.2. Yolov3-Tiny Deep Learning Network

The Yolov3-Tiny model is a simple version of the Yolov3 model. Yolov3-Tiny reduces the number of convolutional layer networks. The basic structure of Yolov3-Tiny has only seven convolutional layers, and its features are extracted by using a small number of and 3 × 3 convolutional layers. Yolov3-Tiny uses the pooling layer instead of Yolov3′s convolutional layer with a step size of 2 to achieve dimensionality reduction shown in Figure 2. The loss function that Yolov3-Tiny uses is the same as that of Yolov3.

2.3. Yolov4 Deep Learning Network

Yolov4 obtains good detection results on a single GPU, such as 1080 Ti and 2080 Ti, and demonstrates a more favorable overall performance. Besides, Yolov4 is easier to use to obtain a high-accuracy model under the single GPU. The prediction time is similar to that of Yolov3. In our study, the classic algorithm modules frequently used in deep learning models for design improvements were carefully selected and tested, and some modules were improved to realize a fast and accurate detector. The improvements were primarily related to the choice of backbone and the integration of several skills. CSPDarknet-53 was selected as the backbone network of the detector, SPP block [21] was added to expand the acceptance flexibility of the model, and the improved model PANet replaced FPN. As for the Tricks, the detection modules most suitable for Yolov4 and most often used in deep learning were selected, including Mish as the activation function and DropBlock as a regularization method. Furthermore, Yolov4 uses a new data enhancement skill called

Mosaic, which expands data by stitching four images together. Several existing methods, including SAM, PANet, and cross-mini batch normalization, were employed to adapt Yolov4 to training with a single GPU. Overall, the main structure of Yolov4 comprises CSPDarknet-53, SPP, PANet, Yolov3 Head, and Tricks, as displayed in Figure 3.

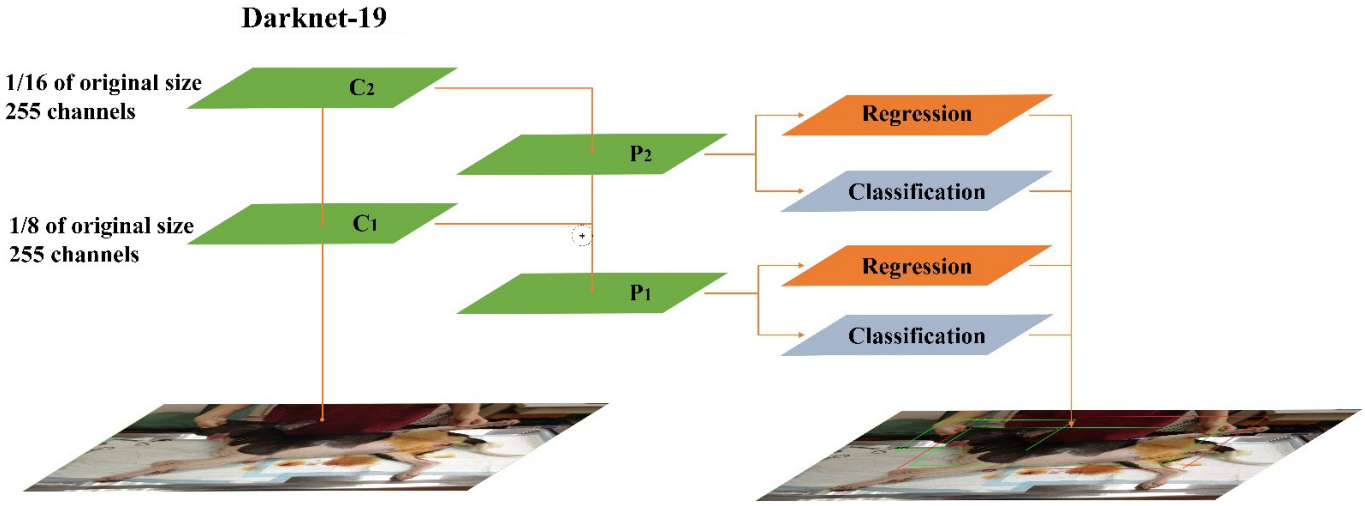


Figure 2. Structure of YOLOv3-Tiny [26] C1,C2 refer to the feature layers of the last two scale. P1,P2 (After superimposition of adjacent layers through upper sampling). ⊕ refers (The superimposition operation).

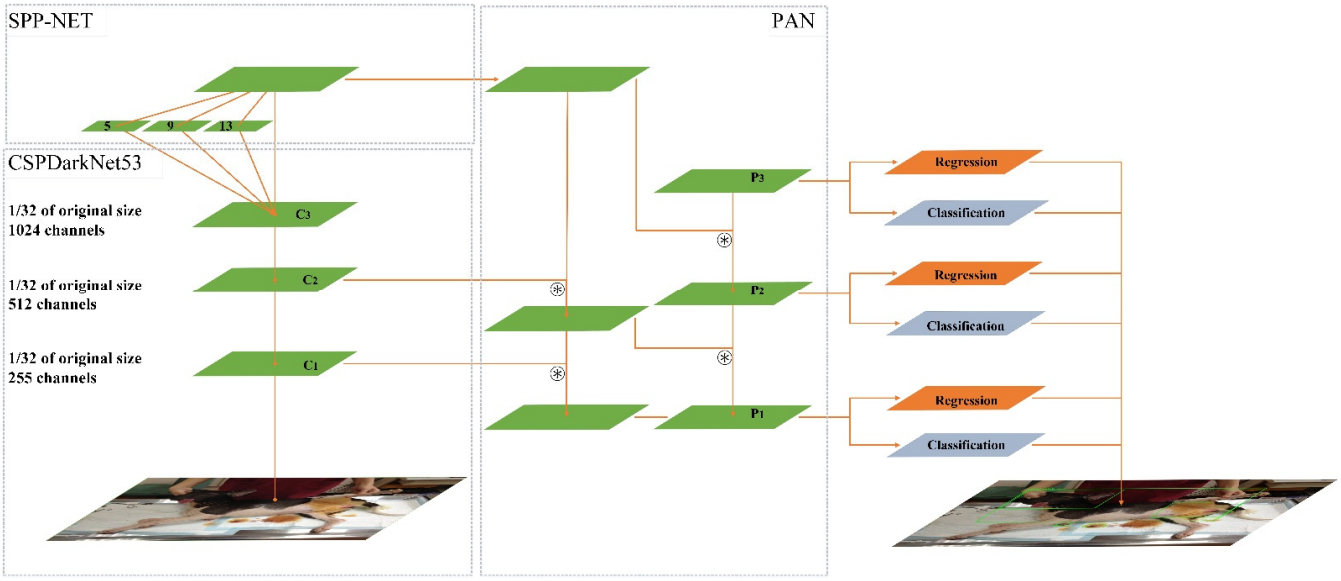


Figure 3. Structure of YOLOv4 [27] C1–C3 refer to the feature layers of the last three scales. SPP refers to spatial pyramid pooling, C3 layer are disposed with 5 × 5, 9 × 9, and 13 × 13 pooling operations respectively.P1,P2,P3 (C1–C3 after aggregation of adjacent layers through down-sampling and up-sampling), ⊛ refers to the aggregation operation.

2.4. Basis of Classifier Voting Technique

The recognition results of the target network may be in different categories, so the target classification needs a post-processing algorithm to make the final judgment and the final classification. The algorithm uses the result of target recognition and then performs a second operation to meet the classification requirements. The method can be an effective method of classifying the TNR dataset. Our post-processing algorithm uses a set vote classifier. This idea can classify the results of the target network, based on which the final decision is taken. Typical approaches are as follows: majority voting (also known as hard

voting), weighted majority voting, and more robust algorithms based on bagging and boosting ideas [46,47], such as a popular AdaBoost.M1 method [48].

$$GP = \{GP_1, \ldots, GP_k\};\ k = 1 \ldots N \tag{3}$$

In the following example, the group GP is a series usually composed of basic classifiers that are not too complicated.

On the assumption that:

Let the decision of the kthclassifier that chooses the jth class be denoted as

$$d_{k,j} \in \{0,1\}; k = 1, \ldots, N; j = 1, \ldots, M, \tag{4}$$

where N is the number of the prediction classifiers and M is the number of predicted classes. If the kth results of object detection choose the jth class, then $d_{k,j} = 1$, and $d_{k,j} = 0$ otherwise. The ensemble mechanism will then choose the featured class j that receives the largest total vote. The situation of weighted majority voting is different from hard voting, as weighted voting defines the factor ω_k according to a certain performance index, which is the weight assigned to the kth classifier, denoted as GP_k.

Here, we define the simplest ensemble voting index DG is as follows.

$$D_G = \underset{j=1,\ldots,M}{argmax} \sum_{k=1}^{N} \omega_k d_{k,j} \tag{5}$$

where k is equivalent to the number of the associated classifier, while j is chosen as the featured class when the pattern recognition is computed. For example, suppose there is a set of three classifiers $GP = \{GP_1, \ldots, GP_K\}$, and three weights ω_1, ω_2, and ω_3, which classify an observation as follows:

$$\begin{array}{l} \text{For } GP_1 \rightarrow \textit{ class } A \text{ and } \omega\_1 = 0.2. \\ \text{For } GP_2 \rightarrow \textit{ class } A \text{ and } \omega\_2 = 0.1. \\ \text{For } GP_3 \rightarrow \textit{ class } B \text{ and } \omega\_3 = 0.5. \end{array}$$

This means that each GP_1, GP_2, and GP_3 classifier chooses the *class A* with ω_1 = 0.2, *class A* with ω_2 = 0.1, and *class B* with ω_3 = 0.5, respectively.

Then, following (5) we calculate the voting index D_G as:

$$\begin{array}{l} \sum_{k=1}^{3} d_{K,A} = 0.2 \times 1 + 0.2 \times 1 + 0.5 \times 0 = 0.4 \\ \sum_{k=1}^{3} d_{K,B} = 0.2 \times 0 + 0.2 \times 0 + 0.5 \times 1 = 0.5 \\ \sum_{k=1}^{3} d_{K,C} = 0.2 \times 0 + 0.2 \times 0 + 0.5 \times 0 = 0 \end{array} \tag{6}$$

Therefore, this maximum voting result will classify the sample as "class B". This algorithm will be used and is very suitable for our application. In this research, a calculated confidence score of detected featured class or category will be used as the weighting factor (ω_k), while one of the four classes, say male dog, will be treated as 1 and 0 will be used for female dog, male cat, and female cat.

3. Methods

The system construction process is described as shown in Figure 4. First, an image recognition framework is created including a training stage and a testing stage (Figure 4). In the training stage, neural network parameters are learned from the training images that have been labeled by a manual, as shown in Figure 5. In the testing stage, the framework that is trained takes incoming images as input and outputs a label prediction. Our developed system is divided into 16 categories. The first four categories are male dogs, female dogs, male cats,

and female cats. Each category has four featured categories including head, body, surgical position, and surgical organs. In the adjusting parameters stage, the result of the classification is then used for majority voting, which will adjust the parameters. Finally, the online stage is implemented.

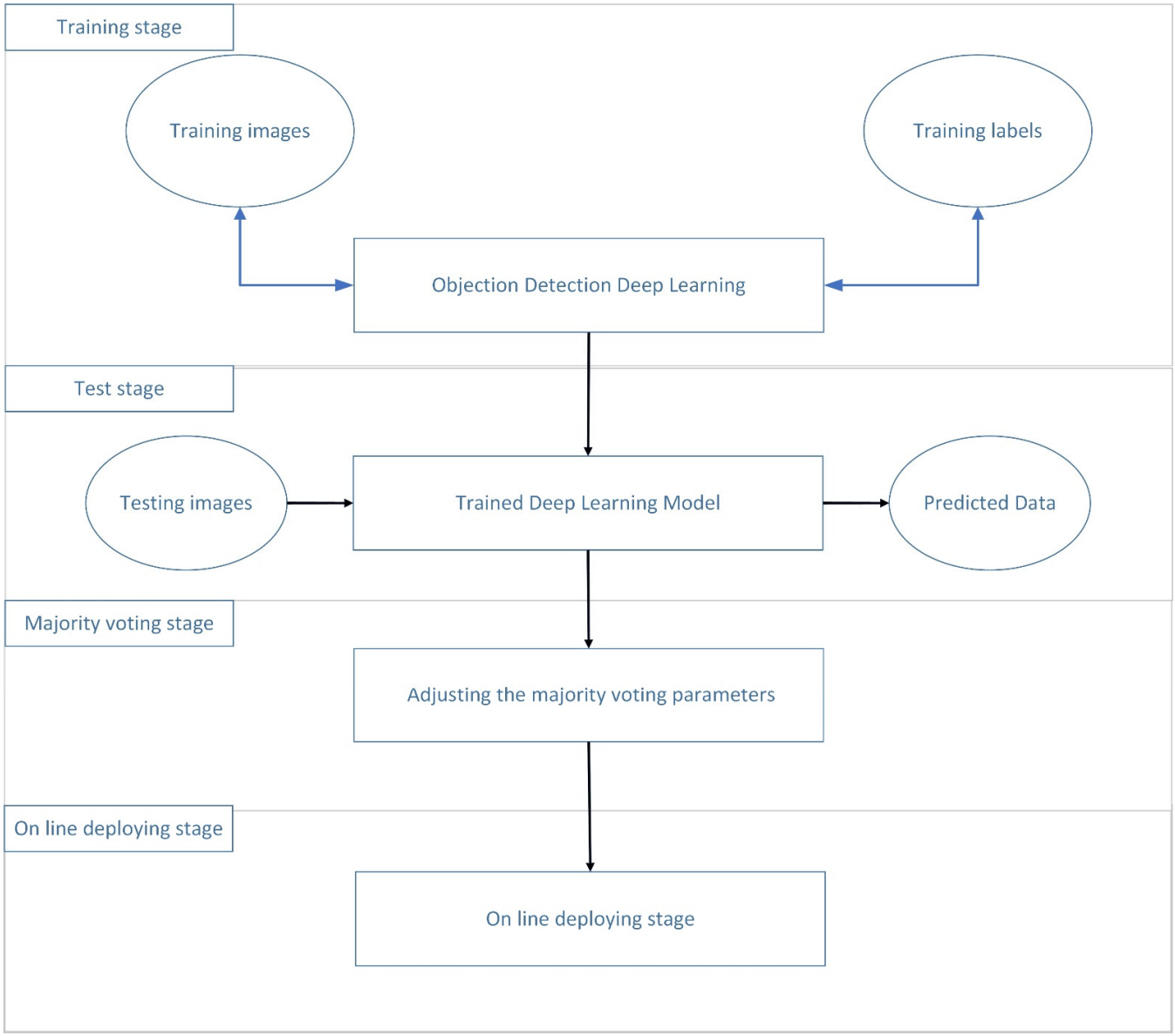


Figure 4. Operation process of Yolo-based feature extraction and majority voting integrated system construction process.

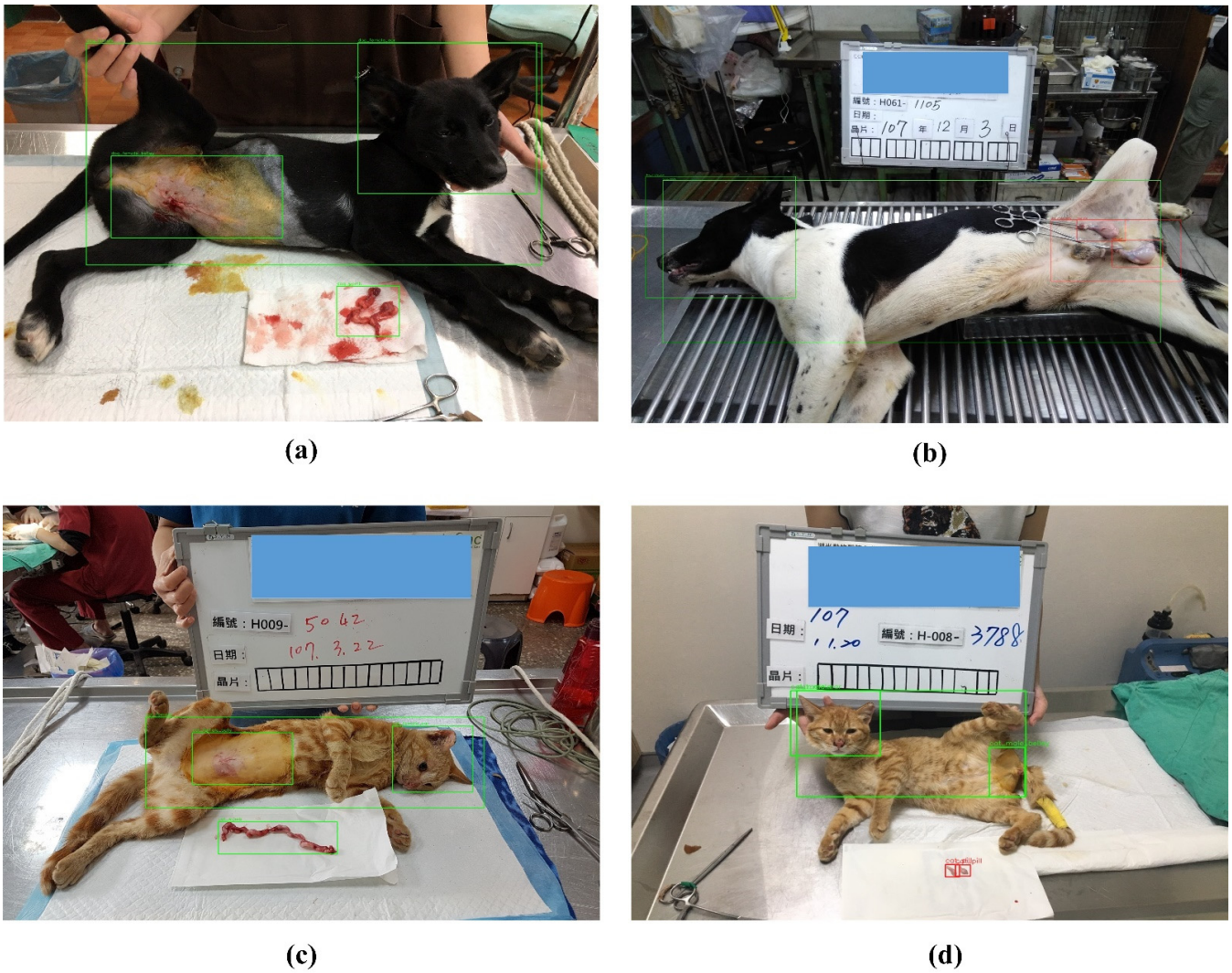


Figure 5. Samples for neutered pictures of Female Dog's (**a**), Male Dog's (**b**), Female cat's (**c**) and Male cat's (**d**) characteristics.

3.1. Processing

The performance evaluations of Yolov3, Yolov3-Tiny, and Yolov4 were conducted in darknet on a GPU PC with the following specifications: Intel Xeon CPU E5-2620 V4 (4.2 GHz × 4) with128 GB of RAM, and Nvidia GeForce RTX 1080 Tis. In the training stage, the momentum of Yolov3 and Yolo3-Tiny was 0.9, the momentum of Yolov4 was 0.949, and bath sizes were set to 64. We trained 50,000 epochs with an initial learning rate of 0.001.

3.2. Dataset Composition and Characteristics

The source of the dataset comes from the HTACT, which is one of the animal protection organizations in Taiwan. These images are used to identify fertility surgery. The images in the dataset are high quality, with resolutions between 960 × 720 and 3000 × 2250 pixels. However, the photos come from different veterinary hospitals, thus standardizing the relevant shooting parameters and shooting angles with the current research is not possible. In total, 813 images were used as the training dataset and 96 images as the testing dataset. The dataset composition is reported and as indicated in Table 2. The training dataset was composed of samples from 73 male dogs, 101 female dogs, 266 male cats, and 373 female cats. The test dataset was composed of samples from 11 male dogs, 19 female dogs, 29 male cats, and 37 female cats. The training dataset accounted for 89% of the total dataset, and the validation dataset accounted for 11% of the total data, as displayed in Table 2.

Table 2. Training and test samples for four classes of TNR.

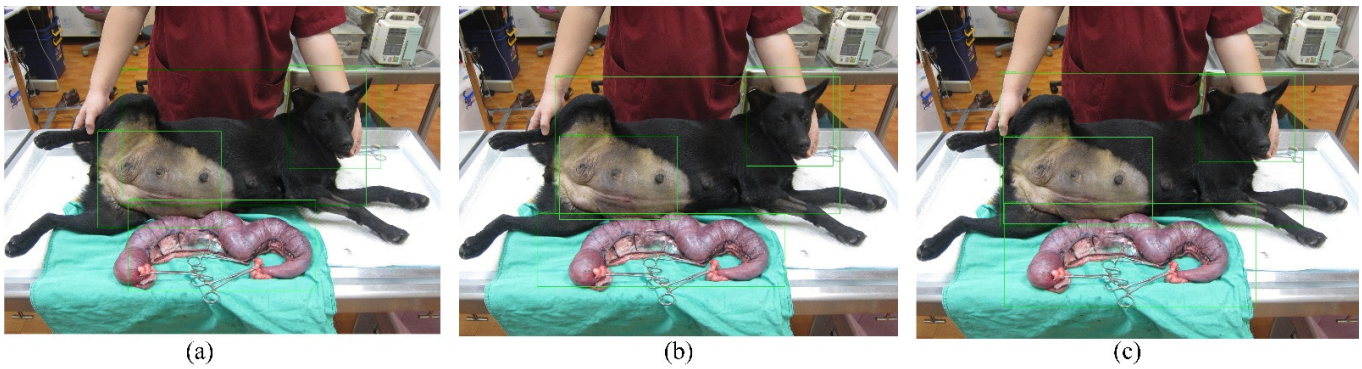
Class	Train	Test
Dog male	73	11
Dog female	101	19
Cat male	266	29
Cat female	373	37
Total	813	96

Figure 5a is a photo of a neutered female dog. The head is facing the right side, and the right ear tip has a cut. The part undergoing surgery is larger than that of the male dog. The removed organ is the female dog's uterus. Figure 5b is a photo of a neutered male dog. The head is facing the left side, and the left ear tip has a cut. The part undergoing surgery is smaller than that of the female dog. The removed organ is the male dog's testis. Figure 5c is a photo of a neutered female cat. The head is facing the right side, and the right ear tip has a cut. The part undergoing surgery is larger than that of the male cat. The removed organ is the female cat's uterus. Figure 5d is a photo of a neutered male cat. The head is facing the left side, and the left ear tip has a cut. The part undergoing surgery is smaller than that of the female cat. The removed organ is the male cat's testis.

The HTACA needs to judge the relative images of each sterilization operation case to see if there are the four corresponding characteristics (categories in this paper) as vouchers. Cats and dogs' ears are cut as a mark that the sterilization operation has been completed, but some cats and dogs with an ear clip after ligation are not photographed, and there are a lot of image records after the operation, as shown in Figure 5.

3.3. Detection Performance

As indicated in Figure 6 and Table 3, the inference results for female dogs are as follows. Yolov3-Tiny (Figure 6a) obtained confidence scores of 0.9994 for the head, 0.9858 for the body, 0.9941 for the surgical position, and 0.9869 for the surgical organs. Yolov3 (Figure 6b) obtained confidence scores of 0.9999 for the head, 1.0000 for the body, 1.0000 for the surgical position, and 0.9959 for the surgical organs. Yolov4 (Figure 6c) obtained confidence scores of 0.9997 for the head, 0.9986 for the body, 0.9926 for the surgical position, and 0.9774 for the surgical organs.



(a) (b) (c)

Figure 6. Pictures of the Yolov3-Tiny (**a**), Yolov3 (**b**) and Yolov4's (**c**) detection for Female Dog.

Table 3. Detection performance for female dog.

	Yolov3-Tiny	Yolov3	Yolov4
Head of female dog	0.9994	0.9999	0.9997
Body of female dog	0.9858	1.0000	0.9986
Surgical position of female dog	0.9941	1.0000	0.9926
Surgical organs of female dog	0.9869	0.9959	0.9774

As displayed in Figure 7 and Table 4, the results of inference for male dogs are as follows. Yolov3-Tiny (Figure 7a) obtained confidence scores of 0.8001 and 0.4046 for the head and body, respectively. Yolov3 (Figure 7b) obtained confidence scores of 0.9948, 0.9993, 0.9995, and 0.9835 for the head, body, surgical position, and surgical organs, respectively. Yolov4 (Figure 7c) obtained confidence scores of 0.9996, 0.9979, 0.9974, and 0.8059/0.9804 for the head, body, surgical position, and surgical organs, respectively. The Yolov4 algorithm has a good ability to recognize small areas, and it can identify the areas of surgical organs of the male dog. The associated two confidence scores are 0.8059 and 0.9804, respectively. This is because Yolov4 can determine two separate images (two balls) from the surgical organs of the male dog, whereas the female dog's organs are connected and lumped together.

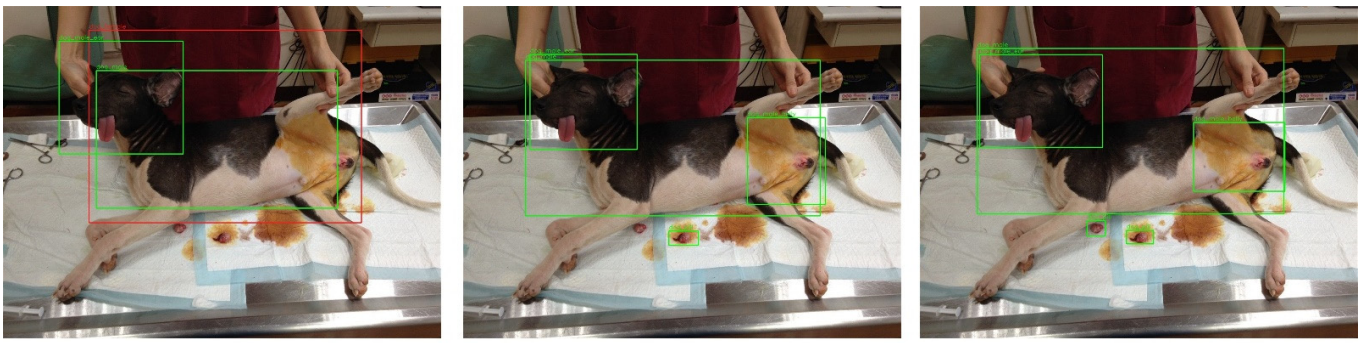


Figure 7. Pictures of the Yolov3-Tiny (**a**), Yolov3 (**b**) and Yolov4′s (**c**) detection performance for Male Dog.

Table 4. Detection performance for male dog.

	Yolov3-Tiny	Yolov3	Yolov4
Head of male dog	0.8001	0.9948	0.9996
Body of male dog	0.4046	0.9993	0.9979
Surgical position of male dog	Null	0.9995	0.9974
Surgical organs of male dog	Null	0.9835	0.8059/0.9804
Body of female dog	0.3417		

The body of the female dog was misjudged by Yolov3-Tiny. Moreover, the confidence score for the body of the male dog was low according to the Yolov3-Tiny, with recognition of only half of the score obtained via the Yolov4, as the architecture for the deep learning model of Yolov3-Tiny is smaller than the Yolov4. These simulations show that Yolov4 performed optimally for small areas and correctly identified the surgical organs of the male dog. Second best in terms of performance was Yolov3. Yolov3-Tiny did not meet the expectations.

As displayed in Figure 8 and Table 5, the inference results for female cats are as follows. Yolov3-Tiny (Figure 8a) obtained confidence scores of 0.9996 for the head, 0.9997 for the body, 0.9914 for the surgical position, and 0.8308 for surgical organs. Yolov3 (Figure 8b) obtained confidence scores of 1.0000 for the head, 0.9997 for the body, 0.9998 for the surgical position, and 0.9961 for surgical organs. Yolov4 (Figure 8c) obtained confidence scores of 0.9991 for the head, 0.9998 for the body, 0.9988 for the surgical position, and 0.8912 and 0.5778 for the surgical organs.

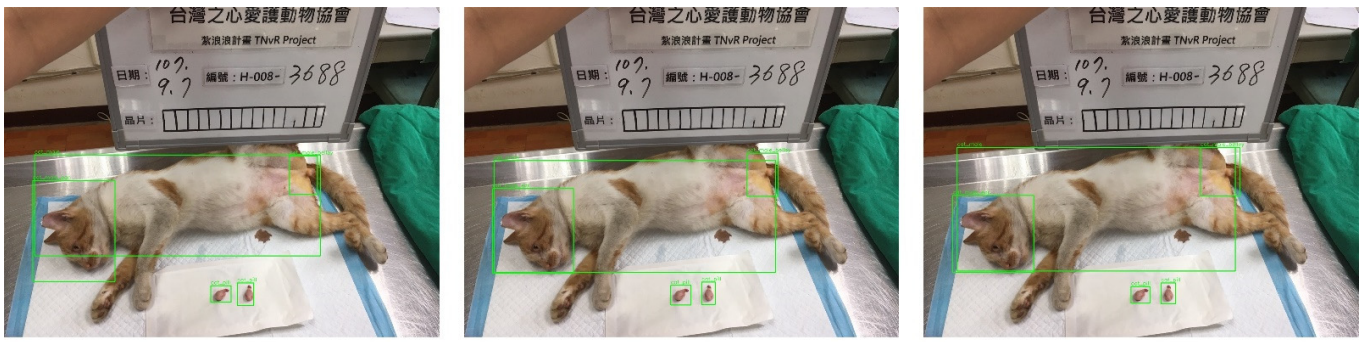
(a) (b) (c)

Figure 8. Pictures of the Yolov3-Tiny (**a**), Yolov3 (**b**) and Yolov4′s (**c**) detection performance for female cat.

Table 5. Detection performance for female cat.

	Yolov3-Tiny	Yolov3	Yolov4
Head of female cat	0.9996	1.0000	0.9991
Body of female cat	0.9997	0.9997	0.9998
Surgical position of female cat	0.9914	0.9998	0.9988
Surgical organs of female cat	0.8308	0.9961	0.8912/0.5778

As displayed in Figure 9 and Table 6, the results of inference for male cats are as follows. Yolov3-Tiny (Figure 9a) obtained confidence scores of 0.9998 for the head, 0.9993 for the body, 0.9539 for the surgical position, and 0.9487 and 0.9135 for the surgical organs. Yolov3 (Figure 9b) obtained confidence scores of 1.0000 for the head, 0.9995 for the body, 1.0000 for the surgical position, and 1.0000 and 0.9999 for the surgical organs. Yolov4 (Figure 9c) obtained confidence scores of 0.9963 for the head, 0.9999 for the body, 0.9961 for the surgical position, and 0.9964 and 0.9966 for the surgical organs. Since the training number of the female cat in Table 2 is 373, the overall confidence score is higher than the confidence score of female dogs, in any case using Yolov3-Tiny, YoloV3, and YoloV4. Nevertheless, complicated structures of YoloV3 and YoloV4 illustrate better confidence scores than the confidence score of Yolov3-Tiny in the categories of body, head, and surgical position of female cats. Still, the confidence score of Yolov3 for the small imaging area of surgical organs of female cats is higher than the score of Yolov3-Tiny. However, Yolov4 can identify the surgical organs of the female cats in two small areas, and their confidence scores are 0.8912 and 0.5778, respectively.



(a) (b) (c)

Figure 9. Pictures of the Yolov3-Tiny (**a**), Yolov3 (**b**) and Yolov4′s (**c**) detection performance for Male Cat.

Table 6. Detection performance for male cat.

	Yolov3-Tiny	Yolov3	Yolov4
Head of male cat	0.9998	1.0000	0.9963
Body of male cat	0.9993	0.9995	0.9999
Surgical position of male cat	0.9539	1.0000	0.9961
Surgical organs of male cat	0.9487/0.9135	1.0000/0.9999	0.9964/0.9966

3.4. Majority Voting Algorithm for Classification Results

Then we will discuss the calculation process and results of deep learning and overall learning Yolo-based majority voting. The object detection classifying models included in this research are Yolov3-Tiny, Yolov3, and Yolov4, which divide the input image into 16 categories (including the 4 categories of male dog, female dog, male cat, and female cat, each of which has another 4 featured characteristics). The total of 16 classification features are classified using a majority voting algorithm to classify male dog, female dog, male cat, and female cat. Four of these classification values are calculated based on the sum of the confidence scores of the corresponding categories of the detection performance, and then compared with the threshold of the majority vote. If the classified value is above the threshold value, then the result of correct TNR surgery can be the output without the need for manual judgment. If the classified value is not greater than the threshold value, then manual judgment is required. According to Table 4, the male dog target classification results were as follows. For Yolov3-Tiny, the sum of the voting index by the male dog confidence score of 0.8001 and 0.4046 was 1.2047, while the voting index for female dog was 0.3417. Hypothetically, if the sum of the four confidence scores for the four categories of the male dog in Table 4 is four times 0.85, as a result, the voting threshold value of male dog was 3.4.

Equation (7) is calculated based on the voting index D_G in Equation (5) as follows.

$$\begin{gathered} \sum_{k=1}^{4} d_{K,Dogmale} = (0.8001 \times 1 + 0.4046 \times 1) = 1.2047 \\ \sum_{k=1}^{4} d_{K,Dogfemale} = 0.3417 \end{gathered} \tag{7}$$

Since the *class Dogmale* = 1.2047 and $class\ Dogfemale = 0.3417$, the $class\ Dogmale > class\ Dogfemale$, and the male dog should be voted as the result. However, if the value of the *class Dogmale* (1.2047) is less than the presetting value of the majority voting threshold (D_G =3.4), then the final majority voting judgment result is "other" (not the male dog). If the classifying model is Yolov3, the presetting majority voting threshold value is 3.6. It means that the confidence score of the four characteristics of male dogs has a performance of 0.9 or higher, and then the voting index is calculated as follows.

$$\sum_{k=1}^{4} d_{K,Dogmale} = (0.9948 \times 1 + 0.9993 \times 1 + 0.999 \times 1 + 0.9835 \times 1) = 3.9766 \tag{8}$$

Since the voting index for *class Dogmale*(3.9766) is larger than the majority voting threshold of 3.6, then $classfied\ as\ male\ dog$ will be returned.

From Table 4, if the classifying model is Yolov4 with the majority voting threshold value of 3.6, then the voting index is calculated as follows.

$$\sum_{k=1}^{4} d_{K,Dogmale} = (0.9996 \times 1 + 0.9979 \times 1 + 0.9974 \times 1 + 0.8059 \times 1 + 0.9804 \times 1) = 4.7812 \tag{9}$$

Again, we have

$$\sum_{k=1}^{4} d_{K,Dogfemale} = \sum_{k=1}^{4} d_{K,Dogmale} = \sum_{k=1}^{4} d_{K,Catfemale} = 0 \tag{10}$$

Since the voting index for *class Dogmale* (4.7812) is larger than the majority voting threshold of 3.6, the classfied as *class maleDog* will be returned.

Thus, the classification results of male dog images by the majority voting algorithms of Yolov3 and Yolov4 for male dogs were 3.9766 and 4.7812, respectively, as shown in Table 7. Based on the same threshold value (denoted as x = 3.4), the majority voting in Figure 10 will be the male dog. The classifying model of using Yolov3 and Yolov4 is good for diagnosing the male dog when the confidence score for the head, surgical position, and surgical organs of male dogs are small.

Table 7. Classification results of male dog images by the majority voting algorithm.

	Yolov3-Tiny	**Yolov3**	**Yolov4**
Male dog	1.2047	3.9711	4.96
Female dog	0.3417		
Categories result	"Other"	Male dog	Male dog

According to the object detection results of the female dog in Table 3, the sum confidence values of Yolov3-Tiny, Yolov3, and Yolov4 are 3.9662, 3.9958, and 3.9683, respectively, and then through majority voting, the classification results by the three classifying models are female dog, as shown in the Table 8. According to the object detection results of female cat in Table 5, the sum confidence values of Yolov3-Tiny, Yolov3, and Yolov4 are 3.8215, 3.9556, and 4.4667, respectively, and then through majority voting, the classification results of the three classifying models are female cat, as shown in Table 9. According to the object detection male cat results in Table 6, the sum confidence values of Yolov3-Tiny, Yolov3, and Yolov4 are 4.8152, 4.9994, and 4.9853, respectively, and then through majority voting, the classification results of the three classifying models are male cat, as shown in Table 10.

Table 8. Classification results of female dog images by the majority voting algorithm.

	Yolov3-Tiny	**Yolov3**	**Yolov4**
Female dog	3.9662	3.9958	3.9683
Classification result	Female Dog	Female Dog	Female Dog

Table 9. Classification results of female cat images by majority voting algorithm.

	Yolov3-Tiny	**Yolov3**	**Yolov4**
Female cat	3.8215	3.9956	4.4667
Classification result	Female cat	Female cat	Female cat

Table 10. Classification results of male cat images by majority voting algorithm.

	Yolov3-Tiny	**Yolov3**	**Yolov4**
Male cat	4.8152	4.9994	4.9853
Classification result	Male cat	Male cat	Male cat

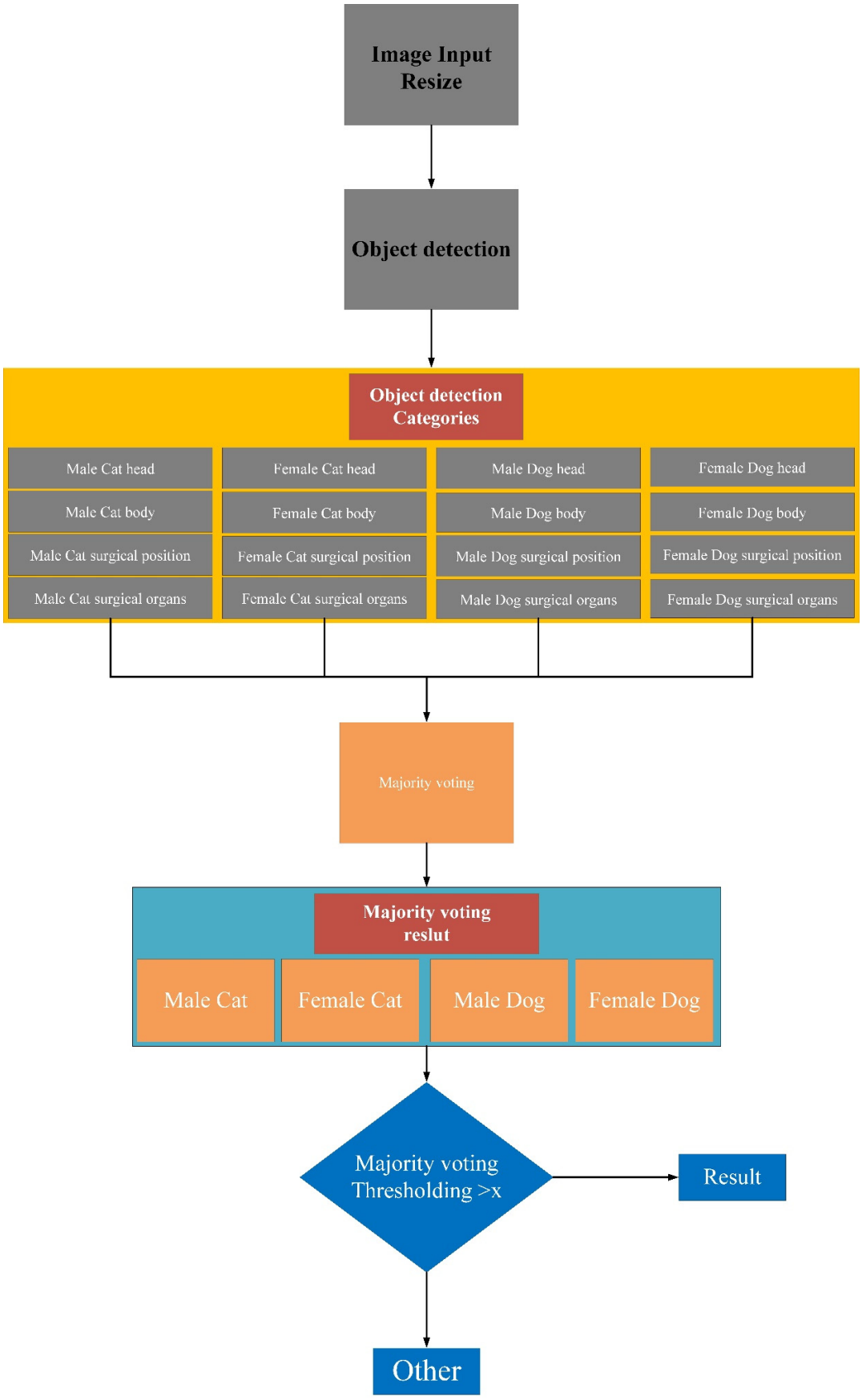


Figure 10. Operation process of Yolo-Based Deep Learning Integrated with Majority Voting System.

4. Discussion

4.1. Evaluation Criteria

TP, *TN*, *FP*, and *FN* represent true positive and true negative and false positive and false negative, respectively.

The calculation of precision in Equation (11) is denoted by the precision for one of images among one of several categories. The calculation of recall is shown in Equation (12).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

The confidence score is obtained based on the probability that a classifier predicts that an object is included in the corresponding bonding box. The IoU for one of images of the one of the classes is the overlap between the prediction box and the original label box (so-called ground truth). The ideal situation is completed overlap, that is, the ratio is 1. Most often, the IoU value is between [0, 1]. Both confidence score and IoU jointly determine the detection result and output whether it is *TP* or *FP*.

The calculation of the average precision (AP) is basically equivalent to calculating the area under the precision-recall curve, which is a two-dimensional curve using accuracy and recall as the vertical and horizontal axes. The performance of the model is evaluated by classification precision, recall, and accuracy. Each category corresponds to an AP. The formula for AP is defined as in Equation (13). As an example, for one head of a male dog from the 16 categories in testing samples, the AP is as follows.

$$\text{Averge Precision}_C = \frac{\sum_{i=1}^{image\ N} \text{Precision}_n \text{ (based on the head of male dog)}}{\text{Total number of the test sample images}} \tag{13}$$

The mean average precision (mAP), the index of measuring detection accuracy in target detection, is the average value of AP values of each category of the multi-category verification dataset, and the corresponding formula is Equation (14).

$$\text{mAP} = \frac{\sum_{c=1}^{c=16} \text{Averge Precision}_c}{\text{Total Category Numbers (16)}} \tag{14}$$

4.2. The Performance of Yolov3-Tiny, Yolov3, and Yolov4

As shown in Figure 11, the mAP of the Yolov3-Tiny and Yolov3 models is 77.68% and 90.92%, respectively; the detection time is 2.68 and 16.52 milliseconds (msec), respectively. The mAP of theYolov4 model is 91.99%, with the detection time of 28.41 msec, as shown in Table 11. Figure 11 Details for the mAP calculation of the Yolov3-Tiny are shown in Table 12. Details for the AP of each category by Yolov3-Tiny and Yolov3 and Yolov4 are as shown in Figures 12–14, respectively.

Table 11. Model efficiency.

	mAP(%)	Detection Time (msec)
Yolov3-Tiny	54.19	2.68
Yolov3	93.99	16.52
Yolov4	62.8	28.41

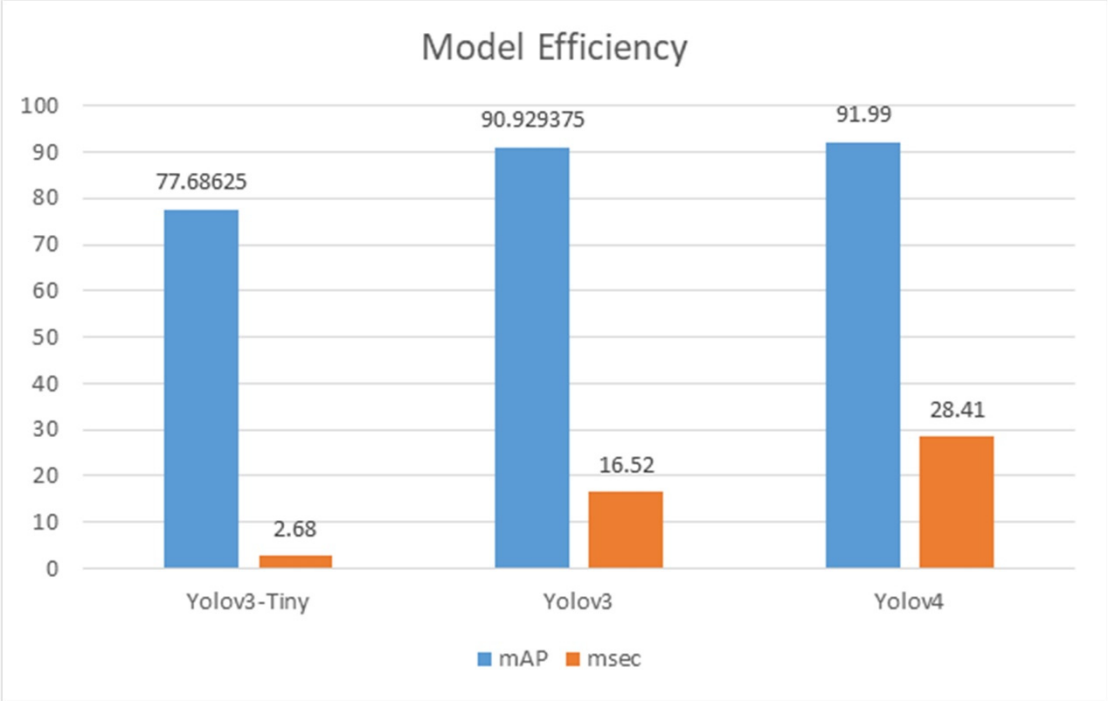


Figure 11. Model efficiency mAP(%).

Table 12. Yolov3-Tiny AP/mAP.

Categories	AP(%)
Head of female dogs	54.19
Body of female dogs	88.43
Surgical position of female dogs	76.92
Surgical organs of female dogs	77.28
Head of male dogs	60
Body of male dogs	63.75
Surgical position of male dogs	60
Surgical organs male dogs	31.82
Head of female cats	93.99
Body of female cats	100
Surgical position of female cats	100
Surgical organs of female cats	78.72
Head of male cats	95.08
Body of male cats	100
Surgical position of male cats	100
Surgical organs male cats	62.8
mAP (%)	77.68

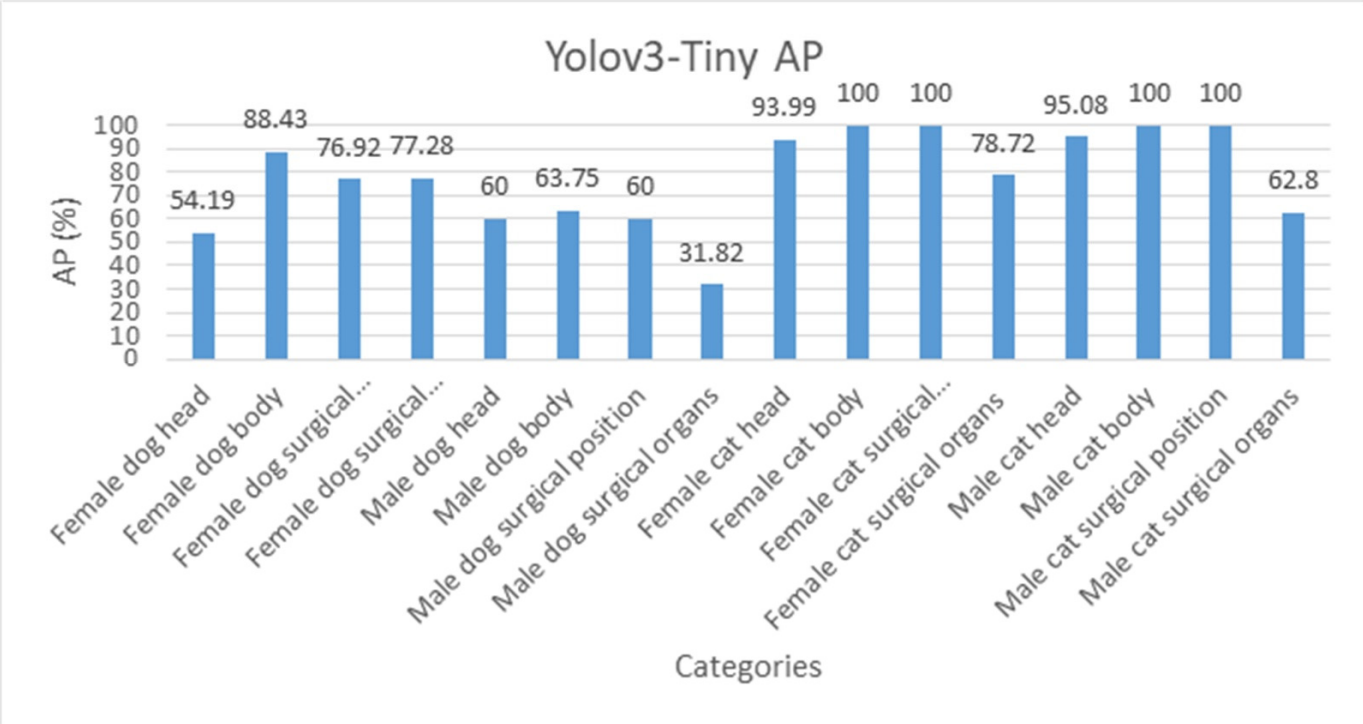


Figure 12. Yolov3-Tiny AP of each category.

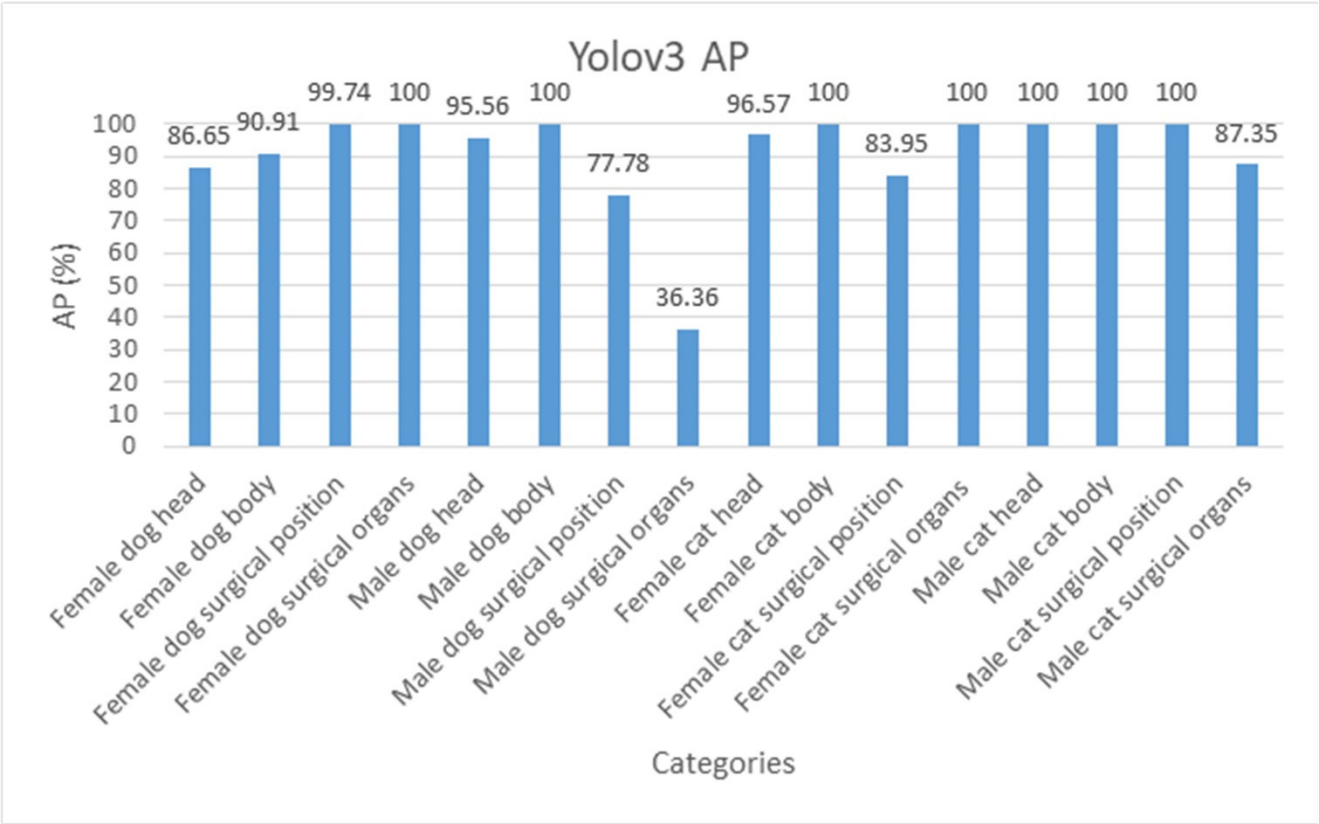


Figure 13. Yolov3 AP of each category.

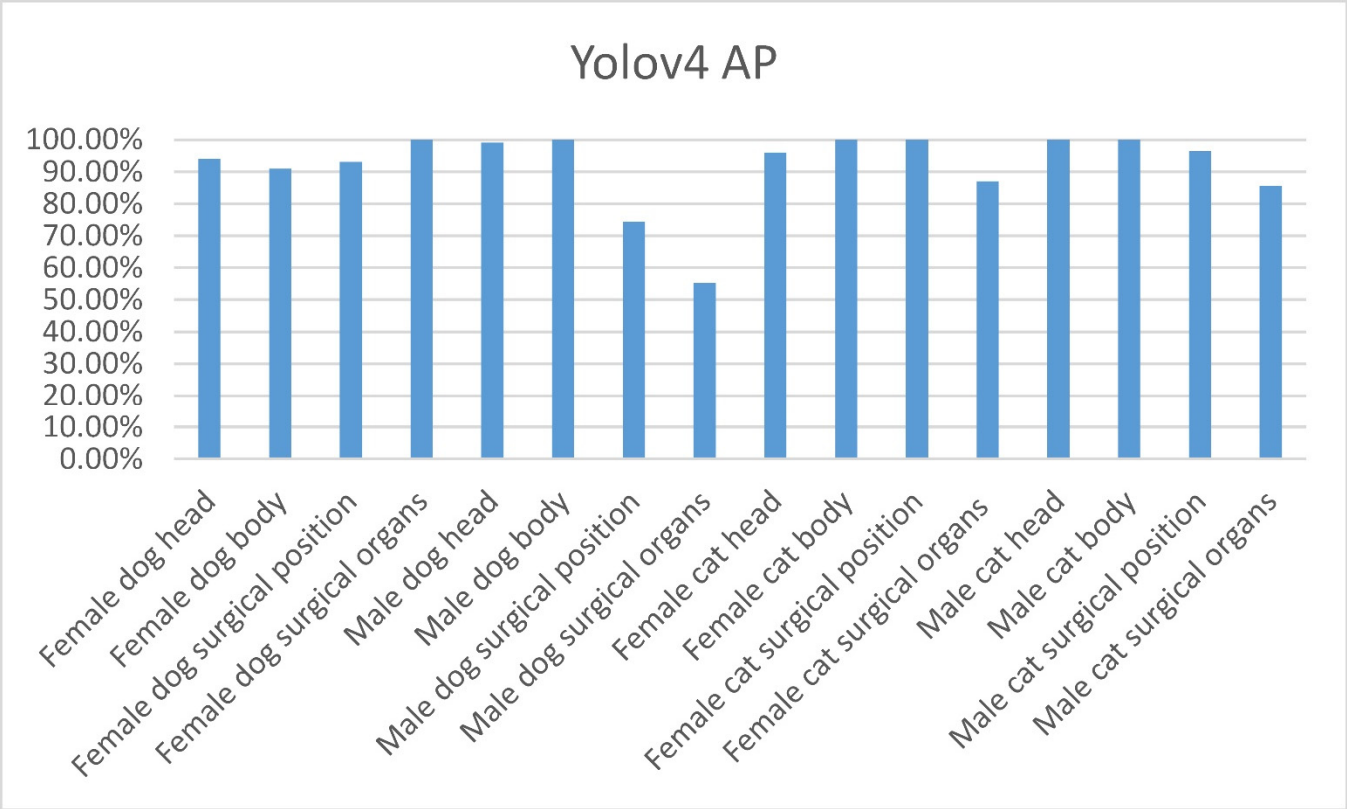


Figure 14. Yolov4 AP of each category.

4.3. The Performance and Confidence Thresholding of Majority Voting Multi-Classification by Yolov3-Tiny, Yolov3, and Yolov4

The mean accuracy is obtained by averaging the accuracy of each class. We use average accuracy because its measurement deviation is smaller than the accuracy in the case of unbalanced datasets. In the training phase, when using an unbalanced dataset, the weights may be biased toward larger groups, so random test images are more likely to be assigned to larger groups. Therefore, in this study, 96 test samples containing 11 male dogs, 19 female dogs, 29 male cats, and 37 female cats were deployed for experimental validation for the proposed model.

In Tables 13–15, the accuracy is the ratio of all predicted correct samples to the total samples. Taking Table 13 as an example, the sum of 70 samples judged to be correct by the majority voting system are 2 male dogs, 7 female dogs, 27 male cats, and 34 female cats. With the "other" by 9, 12, 2, and 3 samples, respectively, there are 96 total samples. Thus, 0.73 is obtained by the ratio of 70 to 96. Individual accuracy is calculated by the ratio of the correct test samples to the total sample number in that specific class. An example for the male dog class, 2 divided by 11 equals 0.18. Mean accuracy is the average of the total sum of each individual accuracy. An example for the mean accuracy (=0.6) of the male dog class is the sum of 0.18, 0.37, 0.93, and 0.92, followed by the division of 4. In Figures 15 and 16, the threshold value parameter x is set as the majority voting value, which is used to determine whether to give a classification category; if the majority voting operation is lower than x, then the output is "other". The number of manual test samples is the number judged as "other" by the majority voting algorithm. More details will be discussed in the Results Section.

Table 13. The performance of Yolov3-Tiny majority voting for multi-classification.

	Male Dog	Female Dog	Male Cat	Female Cat	Accuracy	Mean Accuracy
Male dog	2				0.73	0.6
Female dog		7				
Male cat			27			
Female cat				34		
"Other"	9	12	2	3		
Individual accuracy	0.18	0.37	0.93	0.92		

Table 14. The performance of Yolov3 majority voting for multi-classification.

	Male Dog	Female Dog	Male Cat	Female Cat	Accuracy	Mean Accuracy
Male dog	4				0.802	0.6875
Female dog		8				
Male cat			28			
Female cat				37		
"Other"	7	11	1			
Individual accuracy	0.36	0.42	0.97	1		

Table 15. The performance of Yolov4 majority voting for multi-classification.

	Male Dog	Female Dog	Male Cat	Female Cat	Accuracy	Mean Accuracy
Male dog	9				0.85	0.795
Female dog	1	7				
Male cat			29			
Female cat				37		
"Other"	1	12				
Individual accuracy	0.81	0.37	1	1		

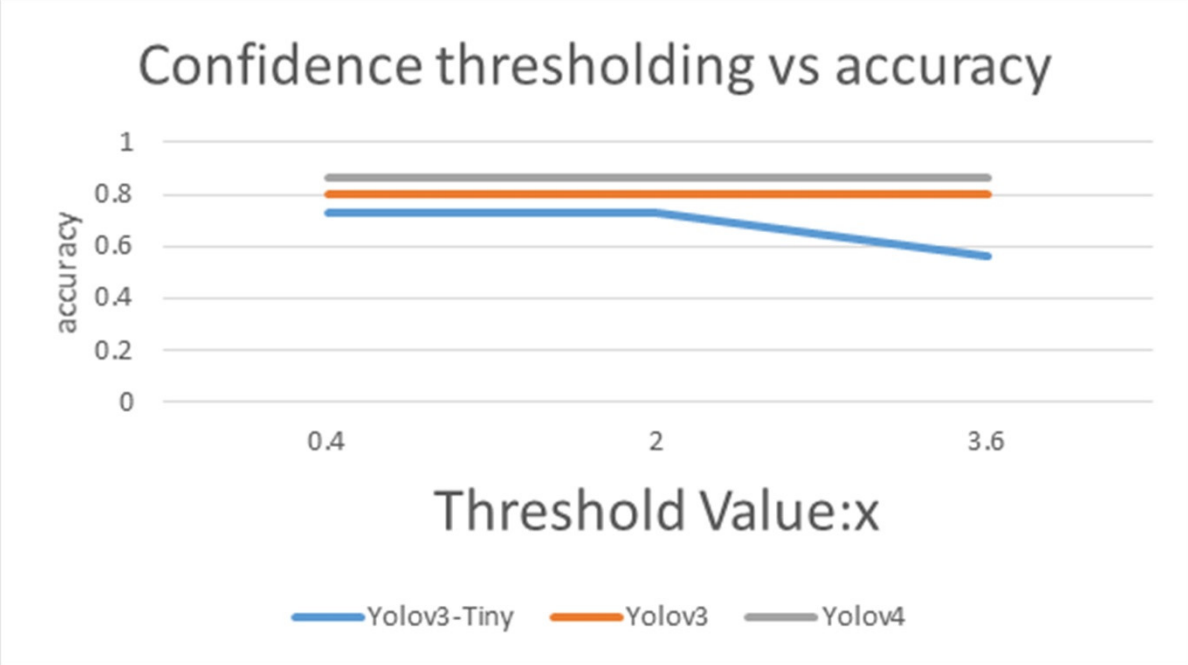


Figure 15. Confidence thresholding vs. accuracy by using Yolov3-Tiny, Yolov3 and Yolov4.

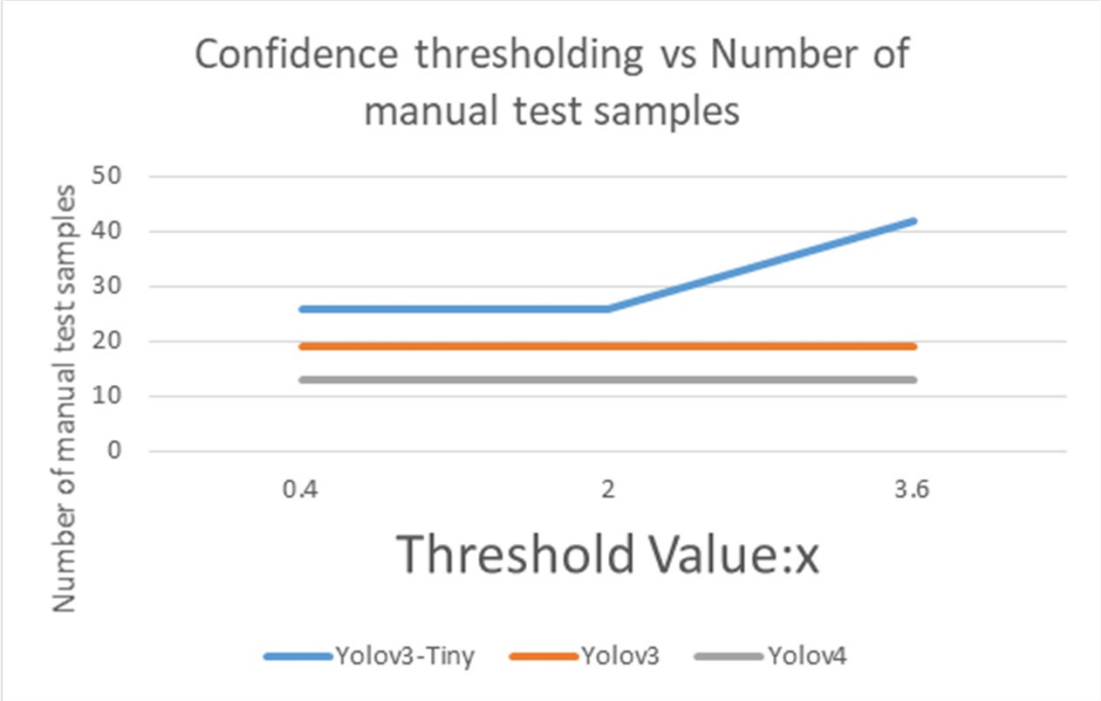


Figure 16. Confidence thresholding vs. Number of manual test samples by using Yolov3-Tiny, Yolov3 and Yolov4.

In Table 13, the accuracy of Yolov3-Tiny majority voting is 0.73, with the mean accuracy of 0.6, when the threshold of Yolov3-Tiny majority voting is set from 0.4 to 2. While the accuracy of Yolov3 majority voting reaches 0.802, and its mean accuracy is 0.687 when the threshold is between 0.4 to 3.6, as shown in the Table 14. As expected, as shown in the Table 15, the accuracy of Yolov4 majority voting is 0.85, with its mean accuracy at 0.795 when the threshold is set from 0.4 to 3.6. As in Figures 15 and 16, when the majority voting threshold of Yolov3-Tiny is set from 0.4 to 2, the accuracy is 0.73, and the number of manual test samples is 26. When the majority voting threshold of Yolov3-Tiny is greater than 2, the accuracy reduces to 0.56, and the number of manual test samples increases to 42. When the majority voting threshold of Yolov3 is set from 0.4 to 3.6, the accuracy is kept at about 0.802, and the number of samples to be manually tested is 19. When the majority voting threshold of Yolov4 is set from 0.4 to 3.6, the accuracy reaches 0.85, and the number of manual test samples is 13. The experimental results found that under the same TNR dataset, Yolov4 works best, Yolov3 takes the second place, and Yolov3-Tiny is less suitable for this dataset.

Some concluding remarks are as follows. If the object detection TNR system judges the wrong category, misclassification will be in FP and FN. This leads to data logging errors and an increase of manpower, administrative process, and cost. For an example, the paid fee for the surgery of female cat is much higher than the male cat. Therefore, for system robustness, an injection of a small percentage of random inspections is recommended. The reliability and stability of the object detection system should be improved. In above experiments, the confidence index of the wrong category was low. Most decision-making algorithms cannot pass the threshold and thus will be judged as "other" here.

4.4. Special Case Discussion

Due to the geographical issue, the number of cats is far more than dogs. It takes times to do the TNR operation. Besides, effective training photo input for neural networks needs professional skill and manual labeling. So, the samples from HTACA's dataset are limited (execution time of this project is from April 2020 until the present). The data balance has a great relevance for the training result of neural networks. To further explore the problem of unbalanced samples, we regrouped a special balanced dataset to conduct the experiment again, as shown in Table 16. We train the neural network using the same hyperparameters.

After 18,000 training epochs, it was found that the Yolov4 neural network reached the mAP of 91.78% on the test dataset, which is very close to the mAP of 91.99% in Yolov4′s training for the original unbalanced dataset. Details for the Yolov4′s AP of each category are shown in Figure 17.

Table 16. Training and test samples for four classes of TNR under the balanced dataset.

Class	Train	Test
Dog male	64	7
Dog female	64	7
Cat male	64	7
Cat female	64	7
Total	256	28

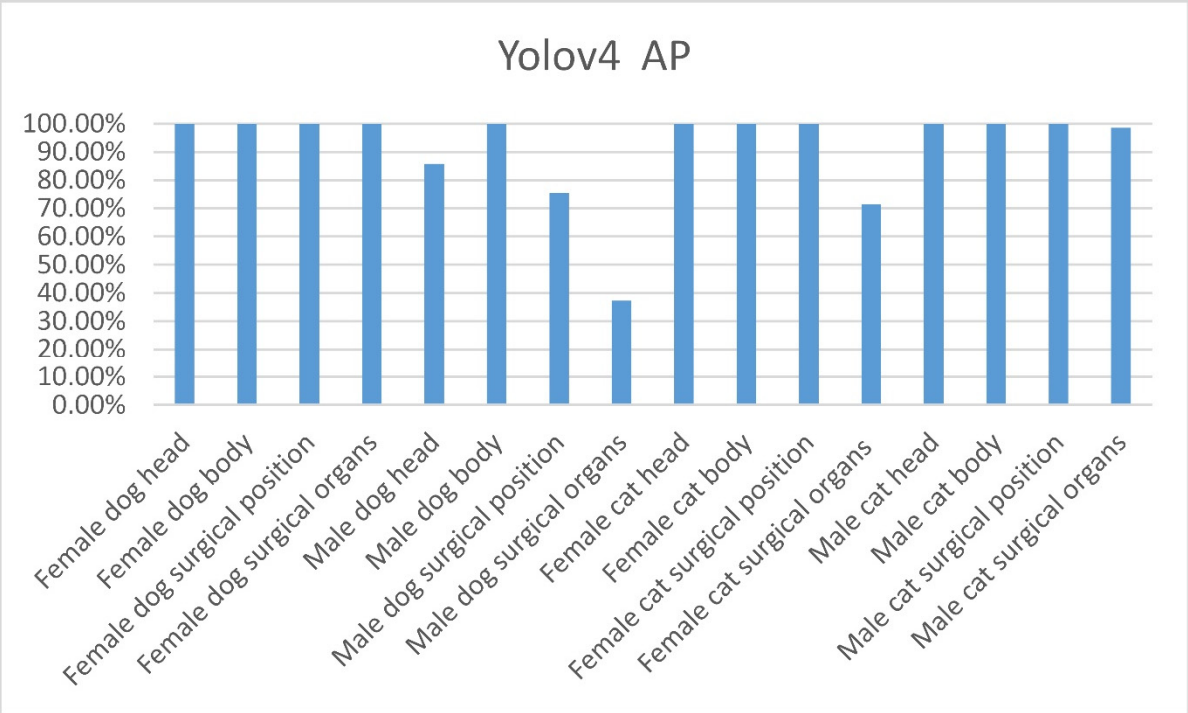


Figure 17. Yolov4 AP of each category under the balanced data set.

As shown in Table 17, when the majority threshold of the balanced dataset is from 0.4 to 3.2, the accuracy and mean accuracy are 0.92, as shown in Figure 18. When the majority vote threshold of Yolov4 is 3.4, the accuracy is reduced to 0.89, and the number of manual test samples is increased to 3, as shown in Figure 19. When the majority vote threshold of Yolov4 is 3.6, the accuracy is reduced to 0.75, and the number of manual test samples is increased to 7. It is stable when the majority threshold of the balanced dataset is 0.4–3.2. From the experiments, it was found that Yolov4 trained on a large dataset is stable compared to models trained on a small dataset. So, the majority vote threshold is relatively easy to be adjusted.

Table 17. The performance of Yolov4 majority voting on multi-classification under the balanced dataset.

	Male Dog	Female Dog	Male Cat	Female Cat	Accuracy	Mean Accuracy
Male dog	5				0.92	0.92
Female dog		7				
Male cat			7			
Female cat				7		
"Other"	2					
Individual accuracy	0.71	1	1	1		

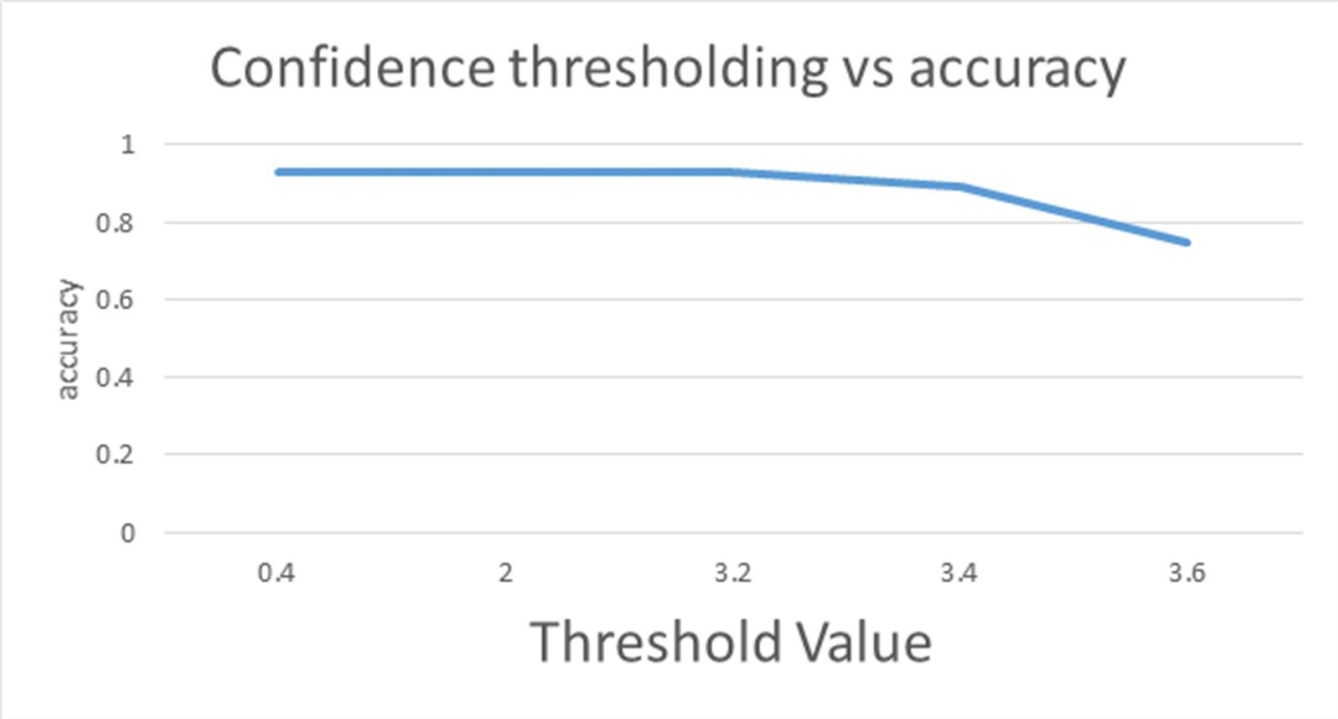


Figure 18. Confidence thresholding vs. accuracy by Yolov4 under the balanced data set.

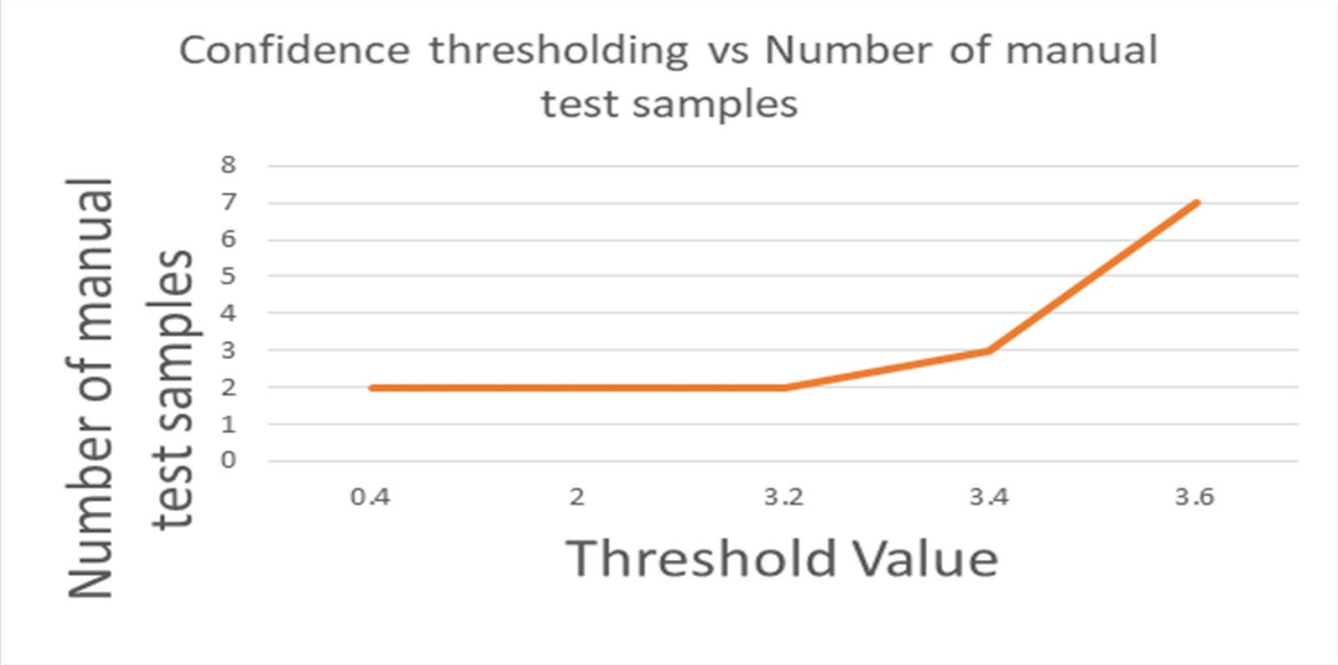


Figure 19. Confidence thresholding vs. Number of manual test samples by using Yolov4 under the balanced data set.

In summary, when a deep learning object detection integrated with majority voting algorithms is used for identifying the TNR sterilization operations through calculation results (unbalanced dataset and special balanced dataset), the trained Yolov4 mAP reaches 91.99%. The result integrates majority voting algorithms' accuracy reaching 85% with an unbalanced dataset. So, it can save more than 85% of manual detection time. Using the best accuracy of the model to evaluate the TNR sterilization can undoubtedly save manual detection time. For object detection with a small number of samples and small areas (such as the recognition of surgical organs of male dogs in images), mAP had the highest value in the experiment. The trained Yolov3 mAP reaches 90.92% and the result that integrates majority voting algorithms' accuracy reaches 80%, so it can save more than 80% of manual detection time. In this research, the performance by Yolov4 is better than the Yolov3. Regarding the Yolov3 and Yolov4′s image classification results, the parameter $\times$ for majority voting threshold (from 0.4 to 3.6) is the least sensitive to the overall classification, so it is the easiest to adjust the majority voting threshold from this research study. Although the mAP of trained Yolov3-Tiny is not high, at only 77.68%, it can save more than 72% of the manual detection time.

5. Conclusions

In this research, Yolov3-Tiny-, Yolov3-, and Yolov4-based feature extraction and majority voting integrated systems used to classify the images of TNR sterilization surgery of neutered animals were proposed and deployed successfully. Experimental results demonstrate that such algorithms can use the spirit of man-machine cooperation for a novel TNR task force. In fact, if the Yolo's model is not accurate enough, the output of the majority voting will be "other". In practice, if the sample is judged as "other", it will go directly to the manual review status. Therefore, a situation caused by misjudgment will not occur. More than 90% of the mAP is reached successfully when the Yolov3 and Yolov4 are implemented. This study shows that the machine learning method with object detection and multi-classifier majority voting can save more than 80% of time for classifying the images of neutered animals. A good effect on the management of stray animals was demonstrated successfully. As the research is continuously being conducted, there will be abundant available input images to be trained in the proposed system and it should be gradually improved with higher accuracy with more labor-saving abilities. Furthermore, the classification category for TNR dataset can be tailored, adjusted, or expanded for images for the purpose of TNR's medical diagnosis. Evaluation for the use of machine learning to automatically adjust the value of the majority voting threshold and the hyper parameters of the loss function based on the recall dataset of manual works will be deliberated in future research. Moreover, the photos taken in different veterinary hospitals are unable to be standardized. The hardware aspects of shooting angles and other related shooting parameters will be considered in a future study. In terms of the Yolo-based DNN efficiency, customization for TNR based on the modification of the CNN backbone and the architecture of feature pyramid networks for object detection can be achieved.

Author Contributions: Conceived, Y.-C.H. and T.-H.C.; wrote the paper, Y.-C.H.; methodology, Y.-C.H., T.-H.C. and Y.-L.L.; software, T.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the Heart of Taiwan Animal Care Association for granting the permission to use the photographs of neutered animals in this research. The authors thank the Ministry of Science and Technology for partially financially supporting this research under Grant MOST 110-2623-E-005-001.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding this publication.

References

1. Patronek, G.J. Free-roaming and feral cats—Their impact on wildlife and human beings. *J. Am. Vet. Med. Assoc.* **1998**, *212*, 218–226. [PubMed]
2. Slater, M.R. *Community Approaches to Feral Cats*; Humane Society Press: Washington, DC, USA, 2002.
3. Levy, J.; Isaza, N.; Scott, K. Effect of high-impact targeted trap-neuter-return and adoption of community cats on cat intake to a shelter. *Vet. J.* **2014**, *201*, 269–274. [CrossRef] [PubMed]
4. Boone, J.D.; Slater, M.A. *A Generalized Population Monitoring Program to Inform the Management of Free-Roaming Cats*; Allian Contracept Cats Dogs: Portland, OR, USA, 2014.
5. Forbush, E.H. *The Domestic Cat; Bird Killer, Mouser and Destroyer of Wild Life; Means of Utilizing and Controlling It*; Wright & Potter Printing Company: Boston, MA, USA, 1916. [CrossRef]
6. Grimm, D. *Citizen Canine: Our Evolving Relationship with Cats and Dogs*; Public Affairs: New York, NY, USA, 2014.
7. Miller, P.S.; Boone, J.D.; Briggs, J.R.; Lawler, D.F.; Levy, J.K.; Nutter, F.B.; Slater, M.; Zawistowski, S. Simulating free-roaming cat population management options in open demographic environments. *PLoS ONE* **2014**, *9*, e113553. [CrossRef]
8. Bennett, P.; Rohlf, V. Perpetration-induced Traumatic Stress in Persons Who Euthanize Nonhuman Animals in Surgeries, Animal Shelters, and Laboratories. *Soc. Anim.* **2005**, *13*, 201–220. [CrossRef] [PubMed]
9. Reeve, C.L.; Rogelberg, S.G.; Spitzmüller, C.; Digiacomo, N. The Caring-Killing Paradox: Euthanasia-Related Strain Among Animal-Shelter Workers1. *J. Appl. Soc. Psychol.* **2005**, *35*, 119–143. [CrossRef]
10. Frommer, S.S.; Arluke, A. Loving Them to Death: Blame-Displacing Strategies of Animal Shelter Workers and Surrenderers. *Soc. Anim.* **1999**, *7*, 1–16. [CrossRef]
11. Baran, B.E.; Allen, J.A.; Rogelberg, S.G.; Spitzmüller, C.; Digiacomo, N.A.; Webb, J.B.; Carter, N.T.; Clark, O.L.; Teeter, L.A.; Walker, A.G. Euthanasia-related strain and coping strategies in animal shelter employees. *J. Am. Vet. Med. Assoc.* **2009**, *235*, 83–88. [CrossRef]
12. Spehar, D.D.; Wolf, P.J. An Examination of an Iconic Trap-Neuter-Return Program: The Newburyport, Massachusetts Case Study. *Animals* **2017**, *7*, 81. [CrossRef] [PubMed]
13. Spehar, D.D.; Wolf, P.J. A case study in citizen science: The effectiveness of a trap-neuter-return program in a Chicago neighborhood. *Animals* **2018**, *8*, 14. [CrossRef] [PubMed]
14. Berkeley, E.P. *TNR Past, Present, and Future*; Alley Cat Allies: Bethesda, MY, USA, 2004.
15. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef]
16. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2009**, *88*, 303–338. [CrossRef]
17. Deng, J.; Karpathy, A.; Ma, S.; Russakovsky, O.; Huang, Z.; Bernstein, M.; Krause, J.; Su, H.; Li, F.-F.; Satheesh, S.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2018**, arXiv:1409.1556.
20. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
22. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Las Condes, Chile, 13–16 December 2015.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inform. Process Syst.* **2015**, *39*, 1137–1149. [CrossRef] [PubMed]
24. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
26. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
27. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
28. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.
29. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
30. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot Multibox Detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016.
31. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

32. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
33. Nguyen, H.; Maclagan, S.J.; Nguyen, T.D.; Nguyen, T.; Flemons, P.; Andrews, K.; Ritchie, E.G.; Phung, D. Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics, Tokyo, Japan, 19–21 October 2017; pp. 40–49. [CrossRef]
34. Villa, A.G.; Salazar, A.; Vargas, F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Inform.* **2017**, *41*, 24–32. [CrossRef]
35. Zeppelzauer, M. Automated detection of elephants in wildlife video. *EURASIP J. Image Video Process.* **2013**, *2013*, 46. [CrossRef]
36. Hsing, P.; Bradley, S.; Kent, V.T.; Hill, R.; Smith, G.; Whittingham, M.J.; Cokill, J.; Crawley, D.; Stephens, P.; MammalWeb Volunteers. Economical crowdsourcing for camera trap image classification. *Remote Sens. Ecol. Conserv.* **2018**, *4*, 361–374. [CrossRef]
37. Delahay, R.J.; Cox, R. Wildlife surveillance using deep learning methods. *Ecol. Evol.* **2019**, *9*, 9453–9466.
38. Spampinato, C.; Farinella, G.M.; Boom, B.; Mezaris, V.; Betke, M.; Fisher, R. Special issue on animal and insect behaviour under-standing in image sequences. *EURASIP J. Image Video Process.* **2015**, 1–4. [CrossRef]
39. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1.
40. Zhu, H.; Yuen, K.-V.; Mihaylova, L.; Leung, H. Overview of Environment Perception for Intelligent Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2584–2601. [CrossRef]
41. Islam, S.S.; Dey, E.K.; Tawhid, M.N.A.; Hossain, B.M. A CNN Based Approach for Garments Texture Design Classification. *Adv. Technol. Innov.* **2017**, *2*, 119.
42. Ho, C.-C.; Su, E.; Li, P.-C.; Bolger, M.J.; Pan, H.-N. Machine Vision and Deep Learning Based Rubber Gasket Defect Detection. *Adv. Technol. Innov.* **2020**, *5*, 76–83. [CrossRef]
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
44. Willi, M.; Pitman, R.T.; Cardoso, A.W.; Locke, C.; Swanson, A.; Boyer, A.; Veldthuis, M.; Fortson, L. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.* **2018**, *10*, 80–91. [CrossRef]
45. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [CrossRef]
46. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; CRC Press: Boca Raton, FL, USA, 2012.
47. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
48. Freung, Y.; Shapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]