mathematics

MDPI

Article

Classification of Fruit Flies by Gender in Images Using Smartphones and the YOLOv4-Tiny Neural Network

Mikhail A. Genaev [1,2,3,4], Evgenii G. Komyshev [1,2,3,4], Olga D. Shishkina [3], Natalya V. Adonyeva [3], Evgenia K. Karpova [3], Nataly E. Gruntenko [3], Lyudmila P. Zakharenko [3], Vasily S. Koval [3,4] and Dmitry A. Afonnikov [1,2,3,4,*]

1 Mathematical Center in Akademgorodok, 630090 Novosibirsk, Russia; mag@bionet.nsc.ru (M.A.G.); delkom07@gmail.com (E.G.K.)
2 Department of Mathematics and Mechanics, Novosibirsk State University, 630090 Novosibirsk, Russia
3 Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia; shishkina.olga.98@gmail.com (O.D.S.); nadon@bionet.nsc.ru (N.V.A.); karpova@bionet.nsc.ru (E.K.K.); nataly@bionet.nsc.ru (N.E.G.); zakharlp@bionet.nsc.ru (L.P.Z.); kovalvs@icg.sbras.ru (V.S.K.)
4 Kurchatov Genomics Center, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia
* Correspondence: ada@bionet.nsc.ru; Tel.: +7-(383)-363-49-63

Abstract: The fruit fly *Drosophila melanogaster* is a classic research object in genetics and systems biology. In the genetic analysis of flies, a routine task is to determine the offspring size and gender ratio in their populations. Currently, these estimates are made manually, which is a very time-consuming process. The counting and gender determination of flies can be automated by using image analysis with deep learning neural networks on mobile devices. We proposed an algorithm based on the YOLOv4-tiny network to identify *Drosophila* flies and determine their gender based on the protocol of taking pictures of insects on a white sheet of paper with a cell phone camera. Three strategies with different types of augmentation were used to train the network. The best performance ($F1 = 0.838$) was achieved using synthetic images with mosaic generation. Females gender determination is worse than that one of males. Among the factors that most strongly influencing the accuracy of fly gender recognition, the fly's position on the paper was the most important. Increased light intensity and higher quality of the device cameras have a positive effect on the recognition accuracy. We implement our method in the FlyCounter Android app for mobile devices, which performs all the image processing steps using the device processors only. The time that the YOLOv4-tiny algorithm takes to process one image is less than 4 s.

Keywords: *Drosophila melanogaster*; gender; image analysis; deep learning; object detection; mobile device; Android app

Citation: Genaev, M.A.; Komyshev, E.G.; Shishkina, O.D.; Adonyeva, N.V.; Karpova, E.K.; Gruntenko, N.E.; Zakharenko, L.P.; Koval, V.S.; Afonnikov, D.A. Classification of Fruit Flies by Gender in Images Using Smartphones and the YOLOv4-Tiny Neural Network. *Mathematics* **2022**, *10*, 295. https://doi.org/10.3390/math10030295

Academic Editor: Radu Tudor Ionescu

Received: 17 December 2021
Accepted: 15 January 2022
Published: 18 January 2022

1. Introduction

1.1. Biological Motivation

Drosophila melanogaster is a classic object for a variety of studies in genetics and systems biology [1]. The evolutionary conservation of the main signaling pathways in the regulation of an animal's metabolism allows the use of *Drosophila* for primary drug testing, which is much faster and cheaper than similar experiments with mammals [2]. One of the traditional ecological indicators in such *D. melanogaster* tests is the offspring size and gender ratio: the genetic effects of drugs are evaluated by the frequency of recessive lethal mutations linked with gender, leading to the selective death of males having only one X-chromosome. This work usually involves counting a large number of offspring in the fly population to assess their fertility and simultaneously determine their gender to estimate their ratio. This task is performed manually and is extremely time-consuming because genetic experiments

require estimating the size of dozens of fly populations, comprising up to several hundred insects [3].

To automate the estimation of the *Drosophila* population size, we previously proposed a smartphone application to obtain and automatically analyze images of flies on a sheet of white paper of standard size [4]. The app is based on computer vision algorithms [5] and allows the user to estimate the number of flies on the sheet with 98% accuracy. These results demonstrate the efficiency of the protocol that is used to acquire images on a cell phone camera: the counting does not require additional equipment or special imaging conditions except for sufficient and uniform illumination. This method, however, does not allow the estimation of the gender of the insects. Thus, the method, which uses the mobile device and image analysis and allows the user to count the number of *Drosophila* and estimate the gender of each fly, is of importance. It should be noted that *D. melanogaster* flies demonstrate sexual dimorphism: not only are females larger than males for most body dimensions, but also the genders differ in pigmentation, the number of visible abdominal segments, the structure of the genitalia, the presence of sex combs, and the shape of various body parts. However, fly gender determination by using a mobile device is complicated because of the small size of the flies (up to 2.5 mm in length).

1.2. Related Works

The automatic identification of insects in digital images, their counting, and species classification are important problems in entomology [6,7] and agriculture [8]. In addition to engineering solutions, computer vision and machine learning are actively used to solve these problems [9]. The most promising results in this area were obtained recently due to the implementation of deep machine learning methods. A number of methods focus on the identification of insects in the field [10–12], which involves distinguishing them from green plants. Some works are aimed at the identification of insects in images from automatic pheromone traps. These images have a homogeneous background, in which color is different from the insects. These types of images are similar to those that are used in our work. The most popular neural network architectures for insect recognition and classification are deep convolutional neural networks (CNNs) [9].

Ding and Taylor [13] distinguished moths among other insects in the trap images with a white background. The CNN was used to solve this problem, and its accuracy was superior to the logistic regression method. The significant improvement in the training of the network was achieved due to the use of data augmentation. Wang et al. [14] analyzed the methods of insect detection in the field images without subsequent classification. Several CNN topologies (VggA, VGG16, Inception V3, ResNet50, CPAFNet, and the model proposed by the authors) were evaluated, and the influence of training parameters on their accuracy and data processing time was investigated. It was shown that the accuracy of the CNN reached 0.91–0.93, depending on the topology and optimization parameters. Liu and Chahl [15] analyzed the CNN algorithms of seven topologies for insect recognition against a natural background. In order to increase the training sample, the authors used the generation of virtual images based on real insect images. For this purpose, real images of insects were segmented; insect contours were rotated randomly and placed on the background image. The use of such synthetic images made it possible to significantly increase the size of the training sample and improve the accuracy of recognition of insects as a result. Note that a similar technique was used to identify barley grains in images using neural networks [16].

A small number of works considered the problem of gender identification for insects. Tuda et al. [17] evaluated a number of machine learning approaches, including logistic regression, random forest, multi-layer perceptron, and support vector machine (SVM), to classify insects of three species (beetle and two parasitic wasps) by gender. Insects were mixed in the images. A total of 2694 features were generated and used for prediction (including shape/size, and color/texture) for each pest image. Authors achieved an

accuracy of 88.5–98.5% for within-species classification of beetles or wasps, 97.3% for two-species classification, and 93.3% for three-species classification.

In Roosjen et al. [18], *Drozophila suzuki* fruit flies were trapped in the field using red, sticky plastic traps. Trap images were acquired in the following two ways: statically, using a digital camera, and dynamically, using a camera mounted on a drone. The authors counted the flies in the images and classified them by gender. The ResNet-18 topology network was applied to image patches on the grid. It was shown that the recall values for female and male identification were 0.73 and 0.68, respectively. However, the area under the curve (AUC) values were 0.506 for females and 0.603 for males, indicating the better performance of the method for the male flies. When gender was not taken into consideration, the recall increased to 0.82 and the AUC to 0.669. The drone images reduced the performance significantly due to lower resolution and non-stationarity.

Recently, the networks based on the YOLO architecture [19] and its modifications have demonstrated rapid identification and classification of objects in digital images. YOLO splits an image into S $\times$ S grids. If the center of an object falls into a grid cell, that grid cell is responsible for detecting the object. YOLO outputs the location of the objects' bounding boxes and their classes on the image along with their confidence. Subsequent works have increased the accuracy of this network architecture and its computational performance, including YOLOv2 [20] and YOLOv3 [21]. The YOLOv4 network has been developed recently [22]. This network proved to be 10% more accurate than YOLOv3 and 12% more computationally efficient. Due to these features, the YOLO network architecture is actively used where efficient data processing is required in the following: in the analysis of images obtained from robot cameras to identify fruits [23], tomatoes [24], and to detect apple flowers in natural environments [25]. One modification of this network, YOLOv4-tiny, was designed to maximize speed and to achieve the lowest computational cost possible [26]. In particular, it has been applied to fruit recognition on drone video [27], plant diseases [28], object tracking [29], and garbage identification based on autonomous trash-collecting robots [30].

In a number of papers, YOLO topology networks have been used to recognize insects in an image. Ramalingam et al. [31] detected and classified insects in indoor and field images. The authors used the Resnet-18 architecture network for recognition, but compared their method to the prediction results of the YOLOv2 network. The recognition accuracy of the YOLOv2 ($F1 = 0.87$) was lower than that of the Resnet-18 ($F1 = 95.79$), while the image processing time was one and a half times lower. Zhong et al. [32] used a Raspberry PI for insect trap image processing using a combination of recognition methods: YOLO network and SVM. The YOLO network was used to detect and coarsely count flying insects, and the SVM was used to classify them. The results demonstrated that the average counting accuracy is 92.50%, the average classifying accuracy is 90.18%, and a cycle of detection and recognition takes about 5 min on a Raspberry PI system. The YOLOv5 network was used for the identification of insects in images of sticky traps located in a eucalyptus forest by Gerovichev et al. [7]. The precision ranged from 0.77 to 0.97 for different types of insects.

Chen et al. [33] proposed a mobile application for insect identification and species classification in the field. It is based on the YOLOv4 network, which showed the highest classification accuracy (100% in mealybugs, 89% in Coccidae, and 97% in Diaspididae) compared to other architectures (region-based CNNs, Faster R-CNNs, and Single Shot Multibox Detectors SSDs). Note that the recognition was performed on the server while the mobile device accessed it via the Internet.

1.3. Contribution of the Work

We implemented the YOLOv4-tiny network for the gender recognition of *Drosophila* flies on the images obtained by mobile device. We have shown that using a learning strategy with synthetic image generation can significantly improve the accuracy of gender recognition in flies. The results of neural network prediction on high-quality images obtained by a digital camera and good illumination were compared with the recognition

performed by expert geneticists. An analysis of the possible sources of recognition errors (lighting conditions, different mobile devices, the gender of the flies, and their position on the paper) was performed. The proposed method was implemented as a FlyCounter mobile application, which has a high speed of image processing.

2. Materials and Methods

2.1. D. melanogaster Lines

To obtain fly images, we used females and males of two laboratory lines of *D. melanogaster* (Harwich and Canton-S) from the collection of the Department of Insect Genetics of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Novosibirsk). Canton-S is one of the most-used wild-type strains in *D. melanogaster* genetics studies [34]. Harwich is a highly inbred wild-type strain of P cytotype [35]. Note that flies of the Harwich line are white-eyed, which distinguishes them in phenotype from Canton-S. Both lines are known as reference lines for the so-called intraspecific paternal-maternal hybrid dysgenesis, which manifests itself in sterility of hybrid offspring in one direction of the cross as a result of nuclear-cytoplasmic interaction [36,37]. Flies were kept on standard food at room temperature and natural light.

2.2. Imaging Protocols

2.2.1. Images for Neural Network Training

When developing our algorithm, it was assumed that the protocol described earlier [4] would be used for counting and classifying flies. In this protocol, flies immobilized with diethyl ether and placed on a white sheet of standard format (A4, A6 or any other from the list provided by the app), with the sheet itself placed against a dark background. Images are taken with a mobile device positioned over the area of the sheet so that the sheet is completely in the frame. The dark background is necessary to recognize white paper and estimate the pixel size/scale of the image.

However, when this protocol is used, there is the problem of labelling the gender of a large number of flies in several hundred images. Accurate identification of the gender requires flipping the fly on the sheet to inspect it from different sides. This requires storing the gender labels of each fly in a separate file, matching them to the location in the image. This procedure proved to be very time-consuming.

Therefore, we applied a modified protocol to organize the learning process of the neural network. Immobilized flies were placed on an A6 sheet as described above. The examiner identified males and females by flipping them on the sheet, then placed the females on the left side of the sheet and the males on the right side, as shown in Figure S1 (Supplementary File). The groups of flies of the same gender on the sheet are well separated in this case. In order to label the gender of groups of insects unambiguously in case of image rotation, a marker (a fragment of white paper) was additionally placed on the male side, which was subsequently used for classifying the groups. Such a protocol makes it possible to avoid additional markers on the sheet to identify the gender of flies; at the same time, the arrangement of flies in each group is as similar as possible to that resulting from the required imaging protocol [4].

The images were taken both with mobile devices and with a Canon EOS 5D Mark IV digital camera. A complete list of devices and the characteristics of their main cameras is given in Table 1. The lighting conditions varied and included bright daylight next to the window, daylight next to the window in cloudy weather, a combination of daylight next to the window and lighting in the room, and daylight lamps. In this way a total number of 365 + 41 = 406 (training + validation) was obtained comprising a total amount of 31,797 + 2073 = 33,870 flies (NET-TRAIN dataset). On average, there were 83 flies per image.

Table 1. List of devices used for fly imaging and characteristics of their main cameras.

Device (Number of Cores/Memory Size)	Main Camera Configuration, Aperture
Xiaomi Mi Max 3 (4/64GB)	12 MP, f/1.90; 5 MP
Xiaomi Mi Note 10 Lite (6/64 GB)	64 MP, f/1.89; 9 MP, f/2.20; 5 MP, f/2.40; 2 MP, f/2.40
Xiaomi Redmi 5 (3/32GB)	12 MP, f/2.20
Samsung Galaxy A3 (SM-A320F)	13 MP, f/1.90
Samsung Galaxy J2 (SM-J250F)	8 MP, f/2.2
Sony Xperia XA	13 MP, f/2.0
Xiaomi Redmi Note 8T (3/32 GB)	48 MP, f/1.75; 8 MP, f/2.20; 2 MP, f/2.40; 2 MP, f/2.40
Xiaomi Redmi Note 9S (4/64 GB)	48 MP, f/1.79; 8 MP, f/2.20; 5 MP, f/2.40; 2 MP, f/2.40
Canon EOS 5D Mark IV	Canon EF 100 mm f/2.8 L lens, aperture 5.0, shutter speed 1/100 sec, ISO 250, manual focus mode

The images vary in size from 5 to 16 Mp (median is 8 Mp), and were downsized during the analysis.

The NET-TRAIN sample obtained in this way was further used to train the parameters of the neural network and evaluate its accuracy (see below).

2.2.2. Protocol for Taking Pictures of Flies on a Grid

In order to evaluate the accuracy of gender determination in flies independently, we used an additional imaging protocol. The immobilized flies were placed on a sheet of white paper in the format of A6. The flies located on a rectangular grid in several rows horizontally and vertically (Figure S2, Supplementary File). The expert classified the flies by gender and described the result for each individual in a legend, which also indicated the imaging device and resolution. The images were taken with mobile devices from the list presented in Table 1. In this way, 42 images were obtained in which a total of 1155 flies were located (NET-TEST dataset). On average, 27 flies were located on one image.

For 23 out of 42 images from the NET-TEST sample the intensity of illumination was estimated with a luxmeter. It was 400 lm for 6 images, 600 lm for 8 images, and 800 lm for 9 images. We used these data to evaluate the effect of illumination on the accuracy of fly classification (see below).

2.3. Labelling Flies in Images

The flies in the images are objects with a complex shape, and can touch each other tightly when placed on the sheet. Therefore, the flies in the image were chosen to be outlined as polygons. The image labelling was performed using the LabelMe program [38] (https://github.com/wkentaro/labelme; accessed at 2 April 2021), see Figure S3 (Supplementary File). This program allows outlining objects of different shapes on the image (rectangles, polygons, circles, lines, dots, line strips) and assigning labels to them. Depending on the position of the fly on the sheet of paper, the outline included the head, body, abdomen, wings, and legs of the fly. The number of vertices of the polygons varied from 10 to 18. Image markup included the file name, image size (width and height in pixels), the list of marked objects (flies), including labels (gender of flies), and coordinates of vertices of polygons. The image labelling information was saved in JSON format.

2.4. Preprocessing Step

At the first stage of the analysis, the identification of the sheet of paper on which the flies were located was performed. This procedure is necessary, on the one hand, to distinguish the white background on which the flies were located, and on the other hand, to determine the scale of the image. The algorithm of paper sheet extraction was described earlier [5] as the following: a paper is recognized as a light area of tetragonal shape on a dark background. For the paper recognition, the original color image is converted to greyscale. To determine the area of the paper, an adaptive binarization of the entire image

is performed. The set of contours is generated and the contour with the largest area is selected. The resulting contour is approximated by a polygon with 4 vertices. If the shape of the paper in the image deviates from rectangular, affine transformation is applied to the image to remove distortion.

2.5. Network Architecture

The YOLOv4-tiny architecture [26] was used to train the fly detection model in the image. Its structure is shown in Figure 1 and was developed based on the YOLOv4 method [22] to provide a higher object detection rate while maintaining object recognition accuracy. In our work the network receives a 1024 × 1024 pixel image as input after selecting a region of a sheet of paper and resizing it.

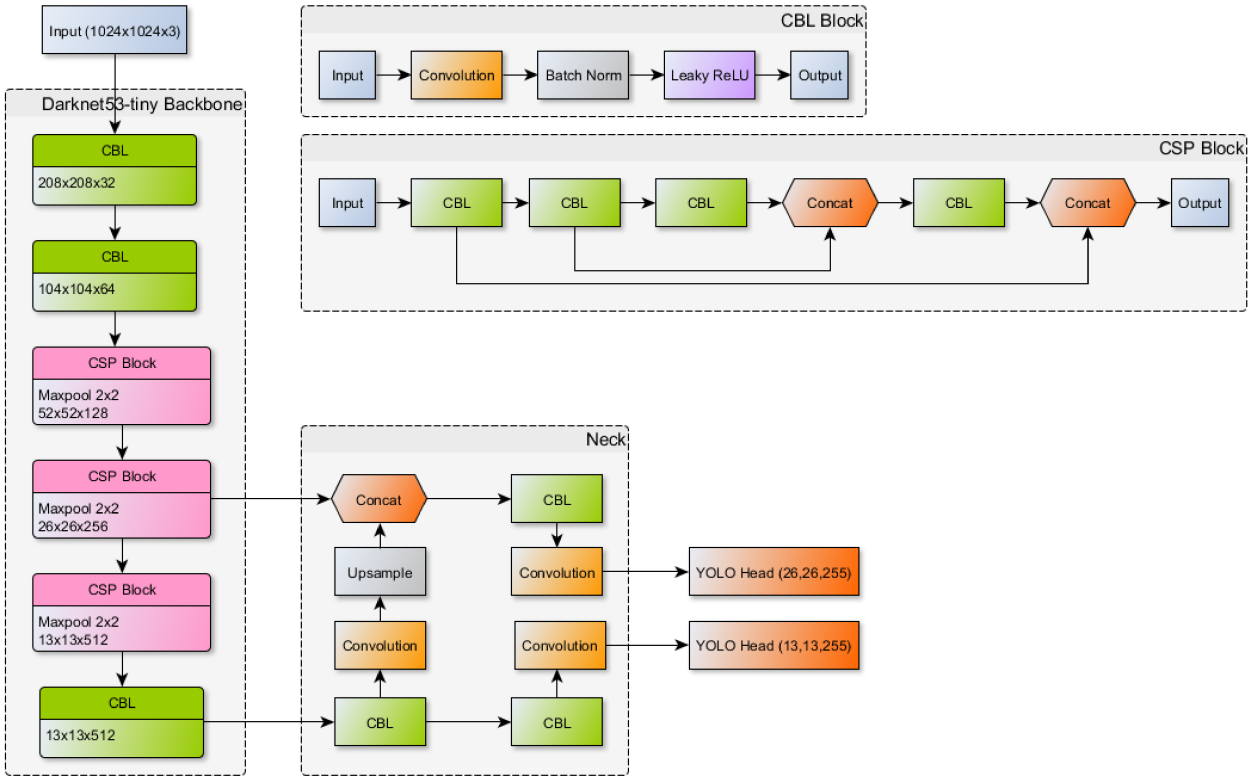

Figure 1. Diagram of the YOLOv4-tiny network architecture used in the work. The main blocks of the network are shown by dashed line rectangles. Block Backbone extracts its features from the input image. The Neck block implements bounding box prediction and object classification based on the extracted features. The three recurring blocks include the following: convolution layers (Convolution, orange), CBL blocks (green), and CSP blocks (pink). The structure of CBL and CSP is shown separately in the diagram.

For feature extraction, YOLOv4-tiny uses CSPDarknet53-tiny as a backbone instead of CSPDarknet53, which is used in the YOLOv4 method. CSPDarknet53-tiny has several differences from CSPDarknet53. CSPDarknet53-tiny uses CBL and CSP blocks for feature extraction (Figure 1) instead of the ResBlock blocks used in CSPDarknet53. The CBL block contains a convolution operation and batch normalization. In addition, CSPDarknet53-tiny uses the Leaky ReLU activation function in the CBL block instead of the Mish function used in CSPDarknet53.

The Leaky ReLU activation feature allows to reduce computational overhead and is defined as follows:

$$y_i = \begin{cases} x_i, x_i \geq 0 \\ x_i/a_i, \ x_i < 0 \end{cases},$$

where a_i is constant parameter which is greater 1.

The CSP block structure uses a feature pyramid network. It ensures that the input feature map, after transformation by CBL, is divided into two parts. The first part remains unchanged, and the second part is convolved, normalized by CBL and divided into two more parts. One of parts remains unchanged, and the second is transformed by CBL. The result is concatenated, transformed by CBL, and then concatenated with the first part of the original data (Figure 1). Such a structure of CSP block allows to reduce computational complexity considerably (by 10–20%), while providing comparable accuracy in object detection.

After processing the data by two consecutive CBS blocks, YOLOv4-tiny divides the input images into grids (grid) of size S × S (S = 26 and 13, see block Neck in Figure 2). For each grid, the network uses three anchors to recognize objects. As a result, S × S × 3 bounding boxes will be created for each input image. The anchors in the grids that contain the centers of the objects will be used to regress the detection boxes.

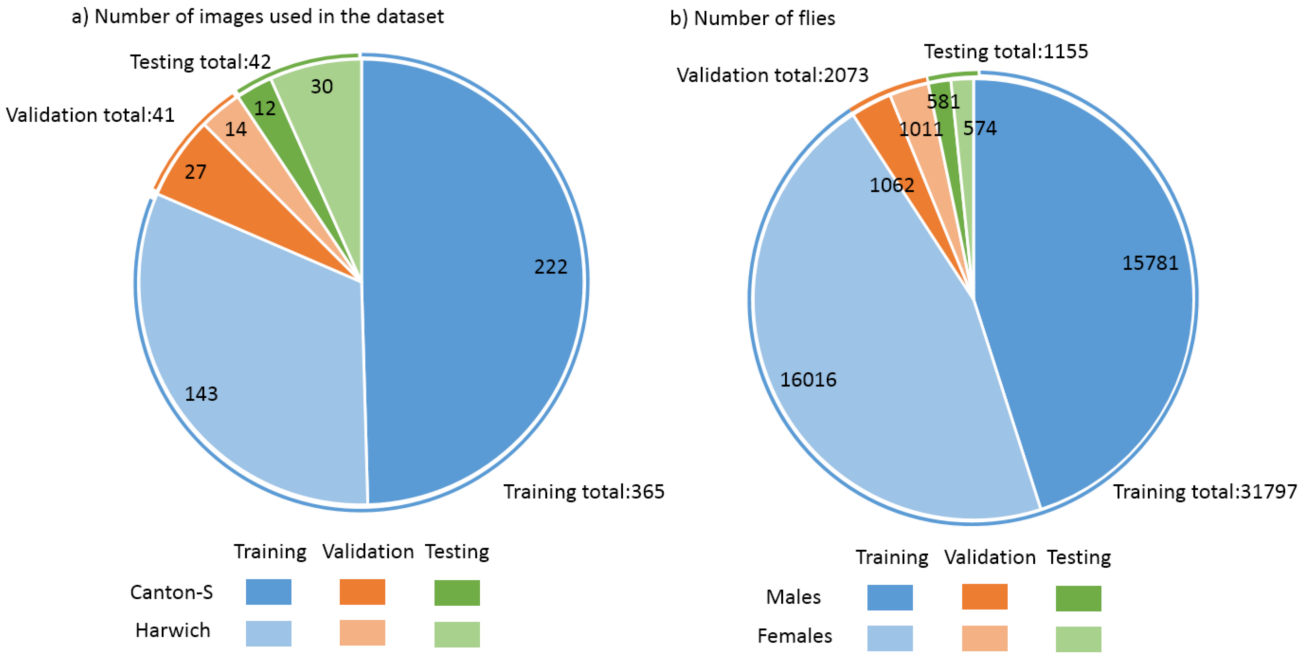

Figure 2. The distribution of the images (**a**) and flies (**b**) within the training, validation, and testing datasets with respect to fly genotype (**a**) and gender (**b**). Bars below the pie diagrams show the color for each dataset class.

To reduce the number of redundant bounding boxes, the confidence of each detection area is calculated. Detections with a confidence level lower than the specified threshold are removed. The detection confidence score of the bounding box j in the i-th grid is defined as follows:

$$Conf_j^i = P_{i,j}(obj) \times IoU_{pred}^{truth},$$

where $P_{i,j}$(obj) = 1 when object is located in the j-th box of the i-th grid; otherwise $P_{i,j}$(obj) = 0. IoU is intersection over union value [39] estimated for predicted and true bounding boxes.

The YOLOv4-tiny loss function is identical to loss function for YOLOv4 and is a sum of the following three values:

$$Loss = loss_1 + loss_2 + loss_3,$$

where $loss_1$ is a bounding box location loss, $loss_2$ is a confidence loss and $loss_3$ is a classification loss.

The YOLOv4-tiny model includes 5,882,634 parameters. The processing performance of a network with such a topology is high: using an NVIDIA 1080Ti GPU, it can reach 371 frames per second. At the same time, the accuracy meets the requirements of a real-world application [26]. As a result, YOLOv4-tiny has significant advantages when solving object detection and classification tasks using mobile devices.

The initial weights obtained from the pre-training of the network on the MS COCO dataset images (https://github.com/AlexeyAB/darknet/releases/download/darknet_yolo_v4_pre/yolov4-tiny.conv.29; accessed on 2 November 2021). The batch size was 64. The model was trained for 6000 iterations. The initial learning rate was learning_rate = 0.001. The learning rate decreased by a factor of 10 when reaching 4800 and 5400 iterations.

2.6. Estimation of the Fly Gender Recognition Performance

To assess the quality of the prediction, after choosing the optimal network parameters, all of the predicted bounding boxes were discarded if the prediction reliability was less than 0.5. The error was estimated from the remaining bounding boxes.

Fly gender recognition was considered true positive (TP) if the *IoU* between the ground-truth and predicted bounding boxes is over 50% and the predicted gender for this fly match its label. Fly gender recognition was considered false positive (FP) if the *IoU* between the ground-truth and predicted bounding boxes is over 50% but the gender is predicted incorrectly. Fly gender recognition was considered false negative (FN) if there are no ground-truth bounding boxes with *IoU* over 50% for the predicted bounding box.

Using these parameters, we estimated precision, recall, and $F1$ measure as follows [40]:

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}.$$

During learning process for each iteration, we additionally evaluated average precision (AP) and mean average precision (mAP) for bounding boxes with *IoU* over 50%, mAP, as described in [41] as follows:

$$AP = \int_0^1 P(R)dR,$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N},$$

where P is precision, R is recall, N is the total number of objects in all categories.

2.7. Synthetic Image Generation

Because a large number of images were required to train the network, we, in addition to the smartphone camera images, used synthetic ones obtained by a method similar to the approach suggested in the work on barley grain image analysis [16]. Synthetic images were generated by combining fly contours and 49 background images of a sheet of paper taken separately with different mobile devices (Galaxy J2, Xiaomi Redmi Max, Xiaomi Redmi 5) under different lighting conditions (daylight, bright daylight, artificial light, daylight lamp). Contours of the flies were obtained by extraction from the original images using the boundaries of the polygons (Figure S3, Supplementary File). For each of these flies, the gender was known. The generation algorithm was as follows: from the available 49 background images, one was randomly selected in an equally probable manner. The number of flies was then determined based on a uniform random distribution between 10 and 90. Image fragments with flies were chosen randomly with equal probability from a total pool of 33,867 fragments (17,024 females and 16,843 males). The arrangement/orientation of the flies on the background was random, without overlapping with previously placed frag-

ments. All synthetic images were generated at a resolution of 1690 × 1200 px. Because the fly contours could have different scales, we reduced the polygons of the flies to the same scale when generating the synthetic image, which was determined based on the pixel resolution of the original image. An example of the synthetic image is shown in Figure S4 (Supplementary File).

2.8. Data Stratification

The total number of images in our dataset is 448. To train and test the *Drosophila* gender recognition algorithm, we divided original images into training, validation, and test samples. Figure 2a shows the distribution of images in the training, validation, and testing datasets with respect to fly genotype. The total number of images of flies used for training/validation/testing is 35025. Figure 2b shows the distribution of the images of flies in the training, validation and testing datasets with respect to fly gender.

Figure 2 demonstrates that our dataset is quite well balanced with respect to fly genotype or gender.

The distribution of the number of images of flies in the training, validation, and test samples and their distribution across different devices is shown in Table 2. Note that for some series of test images, we used devices that were not used in the training sample images (Xiaomi Redmi Note 8T and Xiaomi Redmi Note 9S).

Table 2. Number of images acquired by different devices used for training, validation, and testing the neural network.

Device	Number of Images for Training/Validation/Testing
Xiaomi Mi Max 3 (4/64 GB)	111/9/3
Xiaomi Mi Note 10 Lite (6/64 GB)	8/7/3
Xiaomi Redmi 5 (3/32 GB)	97/9/3
Samsung Galaxy A3 (SM-A320F)	17/0/0
Samsung Galaxy J2 (SM-J250F)	99/7/3
Sony Xperia XA	16/0/0
Xiaomi Redmi Note 8T (3/32 GB)	0/0/23
Xiaomi Redmi Note 9S (4/64 GB)	0/0/7
Canon EOS 5D Mark IV	17/9/0

During training/validation, the dataset was expanded by using synthetic images. The training sample included 2383 images. Of these, 365 are real images with flies from the NET-TRAIN set; 2000 images are synthetic ones. There were a total of 31,797 flies in the real images, including 15,781 males, and 16,016 females. The number of flies per image ranged from 24 to 222. The generated images included fragments with 16,014 females and 15,781 males found in the real images of this subsample. The total number of flies in the synthetic images for the test sample was 99,623 (average of 49.8 flies per image). Additionally, we used 18 images from our previous work [5] with wheat grains as a negative example in training.

The validation sample included a total of 379 images. Of these 41 real images with flies (2073 flies total, 1062 males, and 1011 females), the number of flies per image ranged from 24 to 95. 320 synthetic images were generated from fragments of fly images from this subsample. Additionally, 18 images of wheat grains different from those used for training were included in the validation sample.

The test sample included 42 real NET-TEST dataset images described above (581 males and 574 females). The number of flies per image ranged from 23 to 29. The synthetic images were not used for the test sample.

2.9. Training Strategies

We used three different strategies to train the model. In all cases, the network structure and parameter set were identical (see Section 2.5. Network architecture). The differences consisted of applying different augmentation techniques, using different initial parameter values, and expanding the training sample with synthetic images.

Basic training strategy (YOLOv4-tiny-base). Augmentation involved by random changing image parameters in HSB color space (saturation, brightness (exposure), hue) [42]. For each image, a random change in the three components of this color space was chosen, as described below.

The value of the Hue component (varies between 0–360°) was varied by adding a random value of Hn, chosen from a uniform distribution between $-90°$ and 90°. Saturation (varies from 0–100) was changed by multiplying by either 1/Sn or Sn (chosen with equal probability). Sn was chosen from a uniform distribution between 1 and 1.5. Exposure (varies from 0–100) was varied using a random scaling factor, as it was implemented for Saturation. These changes in HSB components were applied uniformly to all pixels of the image. Additionally, a flip procedure was used, rotating the image by a randomly chosen angle of 90, 270, or 360 degrees.

Strategy using synthetic images (YOLOv4-tiny-synt). The initial weights of this model were equal to the weights of the best model from the YOLOv4-tiny-base strategy. The basic set of augmentations described for the YOLOv4-tiny-base strategy were used. The training and validation samples were supplemented with synthetic images: 2000 images for training, 320 for validation as described above.

Strategy with mosaic generation (YOLOv4-tiny-synt + mosaic). The initial weights of this model were initialized with the weights of the best model from the YOLOv4-tiny-synt strategy. The training and validation samples were supplemented by synthetic images (see description for YOLOv4-tiny-synt strategy). Additionally, images of wheat grains were used as negative data. A basic set of augmentations described for the YOLOv4-tiny-base strategy were implemented. An additional variant of mosaic augmentation was added, where the image is generated based on 4 randomly selected images. The method is taken from Darknet framework library (option mosaic = 1) (https://github.com/AlexeyAB/darknet; accessed on 2 November 2021).

We selected optimal network parameters from the iteration with the maximal *mAP* value (see Section 2.6) for the validation dataset.

2.10. Comparison of Expert Prediction Performance with Network Prediction

In order to compare the accuracy of fly gender recognition by our method with the accuracy of gender identification by experts, we obtained a series of additional images with a Canon EOS 5D Mark IV digital camera (Table 1). Images were taken on a table under studio lighting conditions. We used two Godox 600 sources of direct light with rectangular softboxes 60 × 60 and 60 × 80 cm. The power of sources was 80% of the maximum. The distance from the table with flies to the softbox was 1 m. The camera was mounted on a tripod vertically over an A6 white sheet of paper at a distance sufficient to place it completely in the frame. The resolution of the frame was 5040 × 3360 px. Flies were placed on the sheet according to the protocol used for testing (see previous section). Two genotypes of flies (Canton-S, Harwich) were photographed with 5 images per genotype. There were 29 flies in each image (HUMAN-TEST dataset). These images were not used in training or testing the network.

For the HUMAN-TEST dataset, a preprocessing stage was performed (see Section 2.4. Preprocessing step). The images were printed on a Xerox WorkCentre printer in color mode with a resolution of 1200 dpi. They were provided to the geneticists (co-authors of the paper) to identify the gender of the flies. Performance measures were calculated for each expert's results, as well as for the results of the neural network prediction.

Additionally, we printed in the same way ten images obtained by mobile device and provided them to geneticists for fly gender recognition as described.

2.11. Mobile App FlyCounter for Fly Gender Recognition

To implement the fly gender recognition method on a mobile device, the weights of the best recognition model obtained from the Darknet framework were converted to the TensorFlow format and then converted to the TFLite format using the save_model.py and convert_tflite.py scripts from the repository at https://github.com/hunglc007/tensorflow-yolov4-tflite (accessed on 2 November 2021). The model structure was optimized for fast computation by converting the data from the float32 representation in the original TFLite format to the float16 type.

Based on the obtained model, a mobile application FlyCounter was developed on the Android platform, which counts the number of flies in the image and identifies gender of each fly. The application is implemented on the TensorFlow-lite platform to perform the inference using the mobile device processor. The OpenCV Computer Vision library [43] is used for paper sheet recognition and image preprocessing.

The application works according to the following scheme (Figure 3). The user takes a picture of a sheet of paper with flies. The application performs a perspective correction and crops the paper sheet in the image. The resulting image is fed to the input of the neural network, which outputs the bounding-boxes of the flies in the image and their gender. The application then performs post-processing: it excludes the predicted bounding-boxes with a confidence level < 0.5 and resolves conflicts of overlapping bounding-boxes. As a result, the application displays an image of a sheet of paper with illuminated bounding-boxes corresponding to the predicted flies, as well as the following summary information about the number of flies: total number, number of males, number of females, and the ratio of the number of males to the number of females. The data obtained, the original image and the processed image can be saved to the memory of the mobile device.

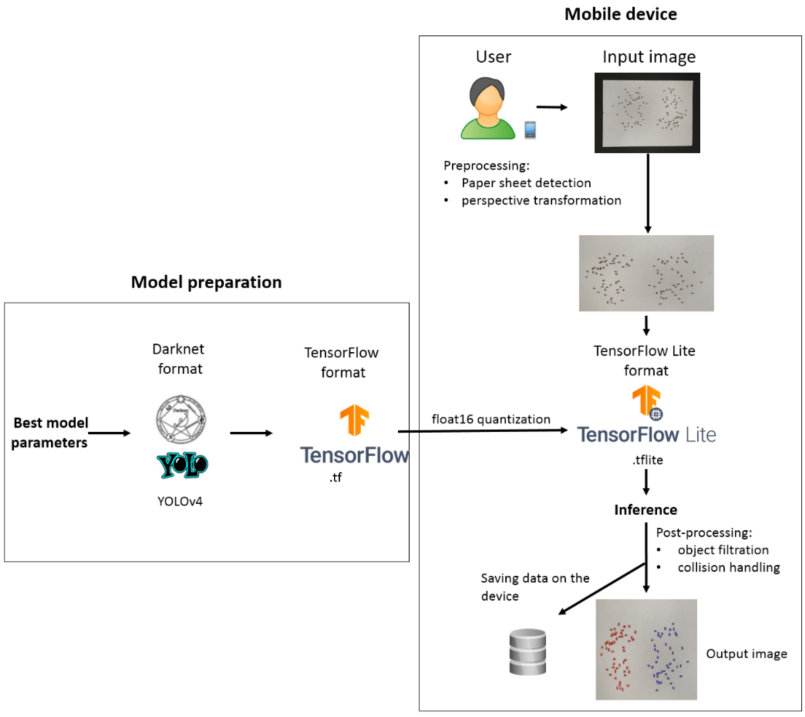

Figure 3. The diagram of the FlyCounter application implementation. The block of the optimal neural network parameters preparation is shown on the left. A user uses the mobile device (right, from top to bottom) to obtain image of the flies on the sheet. The app performs preprocessing, conversion data to tensor-flow lite format, YOLOv4-tiny inference, and outputs the labelled image for the user. The obtained results are stored in the memory of the mobile device.

The app is available at https://play.google.com/store/apps/details?id=ru.delkom.flycounter&hl=en&gl=US (accessed on 16 December 2021).

3. Results

3.1. Fly Gender Recognition Performance by Neural Network

The dependencies of the *Loss* and *mAP* values on the iteration number for the YOLOv4-tiny-synt + mosaic strategy and validation dataset are shown in Figure S5 (Supplementary File). At the beginning of the training process, *Loss* drops quickly during 250 iterations from ~5200 to 40; then it follows a steady decrease with fluctuations. At the end of the training process (iterations 4900 and greater), it approaches two and decreases only slightly. The *mAP* varies between 0.86 and 0.9 during the most part of training. At the end of the training process, it increases to 0.91 and fluctuates around this value.

Table 3 presents the quality assessment metrics for the models implemented with different training strategies on the validation and test (NET-TEST) samples.

Table 3. The results of the fly gender recognition performance evaluation of the different learning strategies on the validation and NET-TEST data samples.

Training Strategy	Validation, Precision	Validation, Recall	Validation, *F*1	Test, Precision	Test, Recall	Test, *F*1
YOLOv4-tiny-base	0.741	0.953	0.834	0.628	0.981	0.766
YOLOv4-tiny-synt	0.852	0.966	0.905	0.700	0.980	0.819
YOLOv4-tiny-synt + mosaic	0.860	0.951	0.904	0.726	0.991	0.838

The precision values consistently increase with more complex learning strategies for both validation and testing datasets. The substantial jump (almost 15% for the validation dataset and 11% for the test dataset) is observed after the expansion of the training/validation datasets by synthetic images with randomized positions of contours of flies. For the best training strategy, the precision values are higher for the validation than for the test dataset (0.860 versus 0.726). At the same time, all proposed models show high recall values on the following validation and test datasets: 0.953 and 0.981 for the YOLOv4-tiny-base, and 0.951 and 0.991 for the best model (YOLOv4-tiny-synt + mosaic). This implies that all models are able to identify the locations of individual flies in the images almost without false positives. The analysis of the model predictions' performance for different types of images showed their advantage as the following: the ability to separate flies from each other when they are densely stacked (touching each other). The expansion of the training dataset by synthetic images had a positive effect on recall for validation data, but not for the test dataset. In general, the use of synthetic images in training significantly improves the classification performance (*F*1 measure, test dataset) by 7.2% relative to the YOLOv4-tiny-base model with a set of basic augmentations on the test dataset. The use of the mosaic method also has a positive effect, raising the recall metric value by 1% relative to the baseline model (YOLOv4-tiny-base). The mosaic generation of the synthetic images (YOLOv4-tiny-synt + mosaic) increases the accuracy by another 2.6% relative to the preceding model (YOLOv4-tiny-synt) on the test dataset. The *F*1 measure of the best model YOLOv4-tiny-synt + mosaic for the test dataset is 0.838, which is lower than for the validation dataset (0.904).

3.2. Comparison of the Performance of Automatic and Expert Recognition

The results of the performance assessment of fly gender recognition by geneticists and the YOLOv4-tiny-synt + mosaic model are shown in Table 4. The average precision of expert classification on a sample of images obtained by mobile devices is 0.716. This is 1% worse than the classification by the YOLOv4-tiny-synt + mosaic model. On images acquired with the Canon 5D Mark IV digital camera and good lighting conditions, the precision of the expert recognition was improved by 18.4% to 0.9. The YOLOv4-tiny-synt + mosaic model is 9.8% better than on the Canon 5D Mark IV images in comparison with cell phone images (precision = 0.824). The average precision of the expert recognition on high-quality images is 7.6% better than the YOLOv4-tiny-synt + mosaic model.

Table 4. Fly gender recognition performance by experts and the YOLOv4-tiny-synt + mosaic network on the independent set of high-quality images.

Prediction Source	Image Source	Precision
Expert 1	Mobile device	0.732
Expert 2	Mobile device	0.713
Expert 2	Mobile device	0.704
YOLOv4-tiny-synt + mosaic	Mobile device	0.726
Expert 1	Canon 5D Mark IV	0.900
Expert 2	Canon 5D Mark IV	0.880
Expert 2	Canon 5D Mark IV	0.920
YOLOv4-tiny-synt + mosaic	Canon 5D Mark IV	0.824

In making this comparison, we estimated the time to manually classify flies on the images obtained from mobile devices (290 flies with 29 flies per image). The average time to label this data was ~11 min. The minimum time was 10 min, the maximum 11 min 47 s. On average, 2.27 s were spent per fly. It is necessary to note that the time required for the exact determination of a fly's gender implementing their flipping is at least three times greater.

3.3. Analysis of Factors Affecting the Accuracy of Recognition

Our results on the NET-TEST test sample were obtained for images of flies on different phone cameras, under different lighting conditions, for two genotypes of flies. It allows us to evaluate the effect of various factors on the accuracy of gender recognition in flies. Table 5 allows us to compare the recognition performance characteristics (TP, FP, FN, and *F*1 measure) for different mobile devices. The *F*1 ranges from 0.696 (Xiaomi Redmi 5 model) to 0.887 (Xiaomi Redmi Note 8T model).

Table 5. Fly gender recognition performance depending on the mobile device on which the images were acquired for the YOLOv4-tiny-synt + mosaic network.

Device	Image Number	TP	FP	FN	*F*1 Measure
Xiaomi Redmi 5	3	39	33	1	0.696
Xiaomi Mi Max 3	3	41	32	0	0.719
Xiaomi Redmi Note 9S	7	123	79	2	0.752
Xiaomi Mi Note 10 Lite	3	47	20	0	0.824
Samsung J2	3	53	20	0	0.841
Xiaomi Redmi Note 8T	23	535	132	4	0.887

It should be noted that the smartphones of different models differ in the number of cameras and their characteristics (Table 1). For instance, the Xiaomi Redmi Note 9S, Xiaomi Mi Note 10 Lite, and Xiaomi Redmi Note 8T devices have 4 cameras each. The Xiaomi Mi Max 3 device has two cameras, while the Xiaomi Redmi 5 and Samsung Galaxy J2 have one camera. We can see that devices with four cameras from the same manufacturer generally outperform devices with fewer cameras. In fact, for Xiaomi devices, the accuracy increases with the increased optics quality (in terms of the number of cameras). The high recognition performance, however, was demonstrated by the Samsung Galaxy J2, which has one camera (an exception to the trend).

Table 6 shows the accuracy estimates obtained with different intensities of illumination of the paper with flies.

Table 6. The performance of fly gender recognition for different paper sheet illumination conditions for the YOLOv4-tiny-synt + mosaic network.

Illumination, lm	Image Number	TP	FP	FN	*F*1 Measure
400	6	134	40	1	0.867
600	8	180	52	2	0.869
800	9	221	40	1	0.915
Without measurement	19	303	184	3	0.764

It can be seen from the table that for 800 lm the *F*1 measure reaches 0.915, which noticeably exceeds the values obtained both at lower illumination conditions (600 and 400 lm) and for the data, for which no illumination measurement was made. Interestingly, the results are not significantly different between the 600 and 400 lm experiments.

Table 7 demonstrates the performance of the fly gender recognition by YOLOv4-tiny-synt + mosaic for two fly lines separately. The *F*1 measures are close, with the difference being less than 1%. Apparently, the difference in fly lines does not significantly affect the accuracy of our method.

Table 7. Performance of fly gender recognition by YOLOv4-tiny-synt + mosaic network for flies from two lines separately.

Line	Image Number	TP	FP	FN	*F*1 Measure
Canton-S	12	222	89	0	0.833
Harwich	30	616	227	7	0.840

Table 8 shows the performance estimation for flies of different genders. The *F*1 measure differs markedly for flies of different genders. It is higher for males compared to females by almost 8%.

Table 8. Performance of fly gender recognition by YOLOv4-tiny-synt + mosaic network for flies of different gender.

Gender	Image Number	TP	FP	FN	*F*1 Measure
Female	574	466	211	4	0.812
Male	581	372	105	3	0.873

3.4. Recognition Performance Analysis Depending on the Position of Flies

We analyzed the effect of the position of flies on a paper sheet on the accuracy of their gender recognition. The analysis was performed on a sample of 10 images obtained using the Canon 5D Mark IV. In this data, most of the errors occur in the recognition of females (43 cases). Misidentification of gender in males was only in three cases. In addition, there were two cases where the fly was not detected in the image.

Manual annotation of 290 flies in these images showed that 83% of the flies laid on the side, 14% on the back, and 3% on the front. The precision of determining the gender of flies on the side was 82% ($F1 = 0.905$), on the back 90% ($F1 = 0.935$), and on the front 100% ($F1 = 1$). Precision is 8% higher when the fly lays on its back than when it is on the side. Note that most flies lie on the side, and it is for this category that recognition accuracy is the lowest.

3.5. FlyCounter Mobile App

The main parts of the interface of our developed mobile application include the following: the main menu, the image acquisition interface, the output screen of the image analysis results, the list of saved results, the screen for viewing the saved results, as well as the interface for the application setup. The interface is shown in Figure 4.

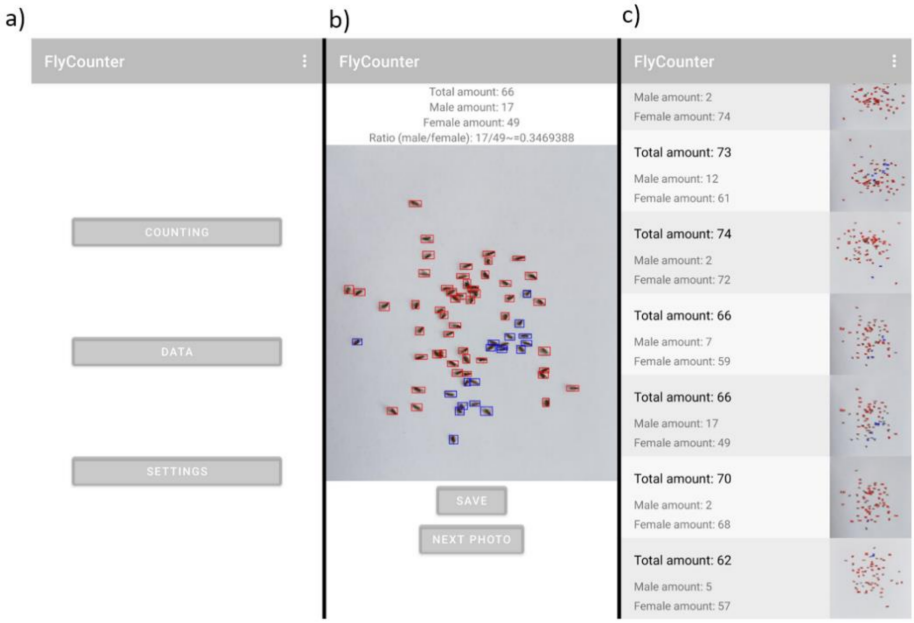

Figure 4. Interface of the FlyCounter mobile application. (**a**). Main menu. (**b**). Image analysis output screen: the summary about the number of females, males, and their ratio is shown at the top of the screen; the center of the screen demonstrates labelled flies. (**c**). List of saved measurements.

At the top of the output screen, a summary of the total number of flies is provided, as well the number of females, males, and their ratio as a decimal fraction. In the center of the screen there is an image of labelled flies (females in red, males in blue). At the bottom of the screen, there are control buttons to save the results and switch to the next image acquisition interface.

We evaluated the data processing time of the YOLOv4-tiny network on mobile devices. The time was measured from the moment of starting its work and the moment of returning the prediction results. We did not include the time taken to obtain an image from the device camera and its preprocessing step. The evaluation was performed for 10 images. The average time per image for neural network data processing does not exceed 4 s: 3.2 s for the Xiaomi Mi Max, 1.4 s for the Xiaomi Mi Note 10 Lite, 3.5 s for the Xiaomi Redmi 5, 3.8 s for the Galaxy A3 SM-A320F, 2.5 s for the Xiaomi Redmi Note 8T, and 1.5 s for Xiaomi Redmi Note 9S. Note that the time for the imaging and preprocessing steps usually takes several seconds. Thus, our app is convenient for instant fly counting and their gender determination. It should be noted that the application does not require Internet access; all data processing is performed on the processor of the mobile device.

4. Discussion

4.1. Choosing the Network Model

Deep learning of neural networks has made significant breakthroughs in the detection and classification of objects in digital images [19,22,26]. Identification and classification of insects in digital images is one of the areas of application for these methods. In such tasks, there are a number of typical problems that have to be solved one way or another to achieve the best result [9].

First of all, it is the practical need to perform data processing on mobile devices. In this regard, exploring opportunities in recognition and classification-based on networks like YOLO is important.

We used the YOLOv4-tiny network topology to recognize gender in images obtained from a cell phone camera. It turned out that the processing of the network data is performed in a few seconds on the processors of mobile devices, which indicates the high computation

performance of this architecture. It demonstrates that similar apps could be developed for insect monitoring in the field, where Internet communication may not be available.

4.2. Dataset Preparation and Expansion by Synthetic Images

Another difficulty with such problems is generating a large, well-annotated dataset of images to train the networks. For such small objects as *Drosophila*, it turns out to be a laborious procedure, taking into account that flies need to be flipped for accurate identification. Despite the rather large size of the sample we obtained, we decided to expand the dataset with the synthetic image generation procedure. This technique has proven useful for solving similar problems [15,16]. We used both the fly contours cropped from the image and the mosaic generation of image blocks in creating such synthetic data. It improved the accuracy of fly recognition by 10% ($F1$ increased from 0.766 to 0.838). Thus, such a technique in the solution of similar problems seems very promising.

4.3. Performance of the Network Model

We showed that the identification of flies in the images, regardless of gender (the recall parameter), is high (0.991). Flies are well-identified, even in the case of their mutual contact in a group of several insects. In general, the identification of touching flies based on neural networks is more reliable compared to the previously proposed application [4], for which the recognition error was about 2%. Thus, the method proposed in this paper allows us to estimate the number of insects in the image with high accuracy.

Our results demonstrate the notable differences between precision/recall/$F1$ values for the validation and testing datasets (Table 3). It should be noted, however, that the location of flies in the images used for training/validation and testing are different. Flies of the same gender in the original training/validation images are located closely on the same part of the paper sheet (Figure S1, Supplementary File). This could result in an unfavorable effect of the CNN training: the classification is affected by nearby flies that have the same gender. In the test, images of the flies of different genders are located randomly on the grid, providing a different local image context for each fly. This could be the reason for the difference in performance metrics for validation and testing datasets. Synthetic images provide random placement of flies in the image irrespectively of their gender and result in a remarkable increase in the performance metrics (YOLOv4-tiny-base versus YOLOv4-tiny-synt strategies, see Table 3).

We were not able, however, to achieve a high degree of accuracy within gender recognition in flies. The precision value of the best model was 73%. This performance can hardly satisfy geneticists when estimating a parameter such as the gender ratio of flies in a population; the error is too high. Thus, the use of our application for solving practical problems related to fly gender identification is not reliable enough. However, this option in our application is present, at least for making a very coarse estimate.

4.4. Comparison with Other Methods and Experts' Evaluation

The high accuracy of gender recognition in insects (beetles and wasps) was shown by the method proposed in Tuda et al. [17] based on SVM. For intraspecific classification by gender, the accuracy was 88.5–98.5%. However, their protocol involved obtaining images of insects on a white background using a scanner. Thus, the quality of the images was sufficiently high.

Gender recognition in *Drosophila* proved to be quite a challenge, not only in our case. Roosjen et al. [18] obtained comparable accuracy characteristics when recognizing *Drosophila* on sticky trap images (AUC was 0.506 for females and 0.603 for males). Note that the imaging conditions in this work were difficult compared to ours. It is interesting, however, that, as in our work, the recognition of males turned out to be more accurate. Based on manual labelling of fly contours, we estimated that the average area projected onto a sheet of paper for males was 3.356 mm^2 (the standard deviation is 0.699 mm^2), and for females, 3.418 mm^2 (standard deviation is 0.683 mm^2). Thus, females are larger

than males by ~1.8%, but the standard deviation of the size of males is larger by 2.2%. Apparently, size alone is not a sufficient factor to classify flies. Probably, males have more pronounced visual manifestations of sexual characteristics (e.g., the presence of a dark spot on the tip of the abdomen), which allows them to be more accurately identified.

A comparison of our method with the experts' evaluation on the same series of images showed that for images obtained with a cell phone camera, the recognition accuracy of the machine algorithm is close to that of the experts. In the case of a high-quality digital camera, high resolution, and good lighting, the experts in the images identify the gender with higher accuracy than our method. This indicates there is an opportunity for improvement in the neural network prediction algorithms.

4.5. Factors Affecting the Performance of the Method

We evaluated the various factors unrelated to the algorithm that affect recognition accuracy. The genotype of flies does not significantly affect the performance, despite the fact that one of the lines (Harwich) differs from the other in the white color of the eyes. Thus, we can judge that our network does retrieve the external traits of flies associated specifically with gender. Based on our analysis, we can conclude that by improving the quality of the image (by increasing its resolution and quality lighting), the recognition accuracy can be improved. However, creating conditions for high-quality imaging will complicate the imaging protocol, which we would like to avoid.

Perhaps the key factor in influencing the accuracy of gender estimation is the position of flies on the paper. Most of them lie on their sides, and it is this position that shows the lowest recognition performance. In a sense, the position on the front or on the back gives, ideal results. However, the position of the flies is the factor that seems to be the least likely to be affected. Apparently, when immobilized, the flies adopt a posture that results in them being more likely to be on their sides than on their back or front.

We can assume several options for improving the protocol by which the recognition accuracy can be improved. First of all, it requires the use of a high-quality digital camera and light sources. However, this would require the allocation of a special workplace for evaluations and the purchase of additional equipment. In addition, a person with special qualifications may be required to set up the lights and camera, and the bright light and additional thermal radiation may adversely affect the viability of the flies.

Our work shows that recognition methods have some difficulties at the current stage, which are mostly caused by imperfections in the protocol of image acquisition (due to its simplicity). However, it is hoped that the improvement of image analysis methods will allow us to achieve better results in the future while maintaining the usability of mobile devices and the speed of data processing.

5. Conclusions

We proposed an algorithm based on the YOLOv4-tiny network to identify and determine the gender of *Drosophila* flies located on a white sheet of paper using a mobile app. Three variants of the training strategy, which differ in the use of synthetic images during training, are investigated. It is shown that training on a sample including synthetic images of flies, generated by superimposing their contours on an artificial background, as well as the mosaic transposition of fragments of the images, allows one to obtain the highest accuracy of recognition. At the same time, the method has a high value of the recall parameter, which indicates a high accuracy in the identification of flies in the image. Gender recognition is less accurate. Among the factors most strongly influencing the accuracy of fly gender recognition, the factor of location on the leaf proved to be the most important. Flies that lie on their sides are recognized as the worst, but their proportion is the highest. In addition, increased light intensity, the higher quality of the device's camera, and increased image resolution had a positive effect on recognition performance. The results also show that the performance of gender recognition is worse in females than in males. The application of YOLOv4-tiny made it possible to implement the fly recognition

method as an application for mobile devices. In this case, the time that the algorithm takes to process one image is less than 4 s.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/math10030295/s1. 'Supplementary File.pdf' (PDF format): Supplementary Figures S1–S5.

Author Contributions: Conceptualization, M.A.G., N.E.G. and D.A.A.; methodology, M.A.G., V.S.K., L.P.Z.; software, M.A.G. and E.G.K.; validation, N.V.A., E.K.K., O.D.S. and L.P.Z.; formal analysis, M.A.G.; investigation, M.A.G.; resources, N.E.G. and D.A.A.; data curation, M.A.G., N.E.G., L.P.Z., O.D.S., E.K.K., V.S.K.; writing—original draft preparation, M.A.G., L.P.Z., N.E.G. and D.A.A.; writing—review and editing, D.A.A.; visualization, M.A.G. and E.G.K.; supervision, N.E.G. and D.A.A.; project administration, D.A.A.; funding acquisition, N.E.G. and D.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: Part of the work (growing of flies, imaging, dataset curation, and labelling) was funded by Ministry of Science and Higher Education of the Russian Federation, project no. FWNR-2022-0019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: Data analysis was performed using the computational resources of the "Bioinformatics" Joint Computational Center of the ICG SB RAS.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AUC	Area under the receiver operation curve
AP	Average precision
CNN	Convolutional neural network
HSV	Color space described by Hue, Saturation, and Value components
IoU	Intersection over union measure
mAP	Mean average precision
SVM	Support Vector machine

References

1. Neumüller, R.A.; Perrimon, N. Where gene discovery turns into systems biology: Genome-scale RNAi screens in *Drosophila*. *Wiley Int. Rev. Syst. Biol. Med.* **2011**, *3*, 471–478. [CrossRef] [PubMed]
2. Pandey, U.B.; Nichols, C.D. Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. *Pharmacol. Rev.* **2011**, *63*, 411–436. [CrossRef] [PubMed]
3. Adonyeva, N.V.; Menshanov, P.N.; Gruntenko, N.A. Link between atmospheric pressure and fertility of *Drosophila* laboratory strains. *Insects* **2021**, *12*, 947. [CrossRef]
4. Karpova, E.K.; Komyshev, E.G.; Genaev, M.A.; Adonyeva, N.V.; Afonnikov, D.A.; Eremina, M.A.; Gruntenko, N.E. Quantifying *Drosophila* adults with the use of a smartphone. *Biol. Open* **2020**, *9*, bio054452. [CrossRef]
5. Komyshev, E.G.; Genaev, M.A.; Afonnikov, D.A. Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front. Plant Sci* **2017**, *7*, 1990. [CrossRef]
6. Høye, T.T.; Ärje, J.; Bjerge, K.; Hansen, O.L.; Iosifidis, A.; Leese, F.; Mann, H.M.R.; Meissner, K.; Melvad, C.; Raitoharju, J. Deep learning and computer vision will transform entomology. *Proc. Nat. Acad. Sci. USA* **2021**, *118*, e2002545117. [CrossRef] [PubMed]
7. Gerovichev, A.; Sadeh, A.; Winter, V.; Bar-Massada, A.; Keasar, T.; Keasar, C. High throughput data acquisition and deep learning for insect ecoinformatics. *Front. Ecol. Evol.* **2021**, *9*, 309. [CrossRef]
8. Cardim Ferreira Lima, M.; Damascena de Almeida Leandro, M.E.; Valero, C.; Pereira Coronel, L.C.; Gonçalves Bazzo, C.O. Automatic detection and monitoring of insect pests—A review. *Agriculture* **2020**, *10*, 161. [CrossRef]
9. Barbedo, J.G.A. Detecting and classifying pests in crops using proximal images and machine learning: A review. *AI* **2020**, *1*, 312–328. [CrossRef]
10. Alves, A.N.; Souza, W.S.; Borges, D.L. Cotton pests classification in field-based images using deep residual networks. *Comp. Electron. Agricult.* **2020**, *174*, 105488. [CrossRef]
11. Ayan, E.; Erbay, H.; Varçın, F. Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks. *Comp. Electron. Agricult.* **2020**, *179*, 105809. [CrossRef]
12. Xie, C.; Zhang, J.; Li, R.; Li, J.; Hong, P.; Xia, J.; Chen, P. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. *Comp. Electron. Agricult.* **2015**, *119*, 123–132. [CrossRef]

13. Ding, W.; Taylor, G. Automatic moth detection from trap images for pest management. *Comp. Electron. Agricult.* **2016**, *123*, 17–28. [CrossRef]
14. Wang, J.; Li, Y.; Feng, H.; Ren, L.; Du, X.; Wu, J. Common pests image recognition based on deep convolutional neural network. *Comp. Electron. Agricult.* **2020**, *179*, 105834. [CrossRef]
15. Liu, H.; Chahl, J.S. Proximal detecting invertebrate pests on crops using a deep residual convolutional neural network trained by virtual images. *Artif. Intell. Agricult.* **2021**, *5*, 13–23. [CrossRef]
16. Toda, Y.; Okura, F.; Ito, J.; Okada, S.; Kinoshita, T.; Tsuji, H.; Saisho, D. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Comm. Biol.* **2020**, *3*, 173. [CrossRef]
17. Tuda, M.; Luna-Maldonado, A.I. Image-based insect species and gender classification by trained supervised machine learning algorithms. *Ecol. Inf.* **2020**, *60*, 101135. [CrossRef]
18. Roosjen, P.P.; Kellenberger, B.; Kooistra, L.; Green, D.R.; Fahrentrapp, J. Deep learning for automated detection of *Drosophila suzukii*: Potential for UAV-based monitoring. *Pest Manag. Sci.* **2020**, *76*, 2994–3002. [CrossRef] [PubMed]
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, real-time object detection. *J. Chem. Eng. Data* **2015**, *27*, 306–308.
20. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
21. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
22. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
23. Kuznetsova, A.; Maleva, T.; Soloviev, V. Using YOLOv3 algorithm with pre- and post-processing for apple detection in fruit harvesting robot. *Agronomy* **2020**, *10*, 1016. [CrossRef]
24. Liu, G.; Nouaze, J.C.; Touko Mbouembe, P.L.; Kim, J.H. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors* **2020**, *20*, 2145. [CrossRef] [PubMed]
25. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comp. Electron. Agricult.* **2020**, *178*, 105742. [CrossRef]
26. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. Scaled-YOLOv4: Scaling cross stage partial network. *arXiv* **2020**, arXiv:2011.08036.
27. Parico, A.I.B.; Ahamed, T. Real time pear fruit detection and counting using YOLOv4 models and deep SORT. *Sensors* **2021**, *21*, 4803. [CrossRef] [PubMed]
28. Li, F.; Liu, Z.; Shen, W.; Wang, Y.; Wang, Y.; Ge, C.; Sun, F.; Lan, P. A remote sensing and airborne edge-computing based detection system for pine wilt disease. *IEEE Access* **2021**, *9*, 66346–66360. [CrossRef]
29. Wu, H.; Du, C.; Ji, Z.; Gao, M.; He, Z. SORT-YM: An algorithm of multi-object tracking with YOLOv4-tiny and motion prediction. *Electronics* **2021**, *10*, 2319. [CrossRef]
30. Kulshreshtha, M.; Chandra, S.S.; Randhawa, P.; Tsaramirsis, G.; Khadidos, A.; Khadidos, A.O. OATCR: Outdoor autonomous trash-collecting robot design using YOLOv4-tiny. *Electronics* **2021**, *10*, 2292. [CrossRef]
31. Ramalingam, B.; Mohan, R.E.; Pookkuttath, S.; Gómez, B.F.; Sairam Borusu, C.S.C.; Wee Teng, T.; Tamilselvam, Y.K. Remote insects trap monitoring system using deep learning framework and IoT. *Sensors* **2020**, *20*, 5280. [CrossRef]
32. Zhong, Y.; Gao, J.; Lei, Q.; Zhou, Y. A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors* **2018**, *18*, 1489. [CrossRef] [PubMed]
33. Chen, J.-W.; Lin, W.-J.; Cheng, H.-J.; Hung, C.-L.; Lin, C.-Y.; Chen, S.-P. A smartphone-based application for scale pest detection using multiple-object detection methods. *Electronics* **2021**, *10*, 372. [CrossRef]
34. Colomb, J.; Brembs, B. Sub-strains of *Drosophila* Canton-S differ markedly in their locomotor behavior. *F1000Research* **2015**, *3*, 176. [CrossRef] [PubMed]
35. Mackay, T.F.C.; Lyman, R.F.; Jackson, M.S. Effects of P-element mutations on quantitative traits in *Drosophila melanogaster*. *Genetics* **1992**, *130*, 315–332. [CrossRef]
36. Konaç, T.; Bozcuk, A.N.; Kence, A. The effect of hybrid dysgenesis on life span of *Drosophila*. *AGE* **1995**, *18*, 19–23. [CrossRef]
37. Zakharenko, L.P.; Petrovskii, D.V.; Dorogova, N.V.; Putilov, A.A. Association between the effects of high temperature on fertility and sleep in female intra-specific hybrids of *Drosophila melanogaster*. *Insects* **2021**, *12*, 336. [CrossRef] [PubMed]
38. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vision.* **2008**, *77*, 157–173. [CrossRef]
39. Everingham, M.; van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
40. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A Survey on performance metrics for object-detection algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.
41. Yu, J.; Zhang, W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors* **2021**, *21*, 3263. [CrossRef]
42. Busin, L.; Vandenbroucke, N.; Macaire, L. Color spaces and image segmentation. *Adv. Imaging Electron Phys.* **2008**, *151*, 65–168.
43. Kaehler, A.; Bradski, G. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016.