Original Article

Towards toxic and narcotic medication detection with rotated object detectors

Jiao Peng[1#], Feifan Wang[2#], Xiaochi Ma[1], Zichen Chen[2], Zhongqiang Fu[2], Yiying Hu[2], Xinghan Zhou[2], Lijun Wang[1]

[1]Department of Pharmacy, Peking University Shenzhen Hospital, Shenzhen, China; [2]Shenzhen NuboMed Technology Co., Ltd., Shenzhen, China

Contributions: (I) Conception and design: J Peng, F Wang; (II) Administrative support: X Zhou, L Wang; (III) Provision of study materials or patients: J Peng, X Ma, L Wang; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: F Wang, J Peng; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Dr. Feifan Wang. Shenzhen NuboMed Technology Co., Ltd., Block B, Building 6, Shenzhen International Innovation Valley, Nanshan District, Shenzhen 518000, China. Email: woodywff@aliyun.com or feifan.wang@nubomed.com.

Background: Recent years have witnessed the advancement of deep learning vision technologies and applications in the medical industry. Intelligent devices for specific medication management could alleviate workload of medical staff by providing assistance services to identify drug specifications and locations.

Methods: In this work, object detectors based on the you only look once (YOLO) algorithm are tailored for toxic and narcotic medication detection tasks in which there are always numerous of arbitrarily oriented small bottles. Specifically, we propose a flexible annotation process that defines a rotated bounding box with a degree ranging from 0° to 90° without worry about the long-short edges. Moreover, a mask-mapping-based non-maximum suppression method has been leveraged to accelerate the post-processing speed and achieve a feasible and efficient medication detector that identifies arbitrarily oriented bounding boxes.

Results: Extensive experiments have demonstrated that rotated YOLO detectors are highly suitable for identifying densely arranged drugs. Six thousand synthetic data and 523 hospital collected images have been taken for training of the network. The mean average precision of the proposed network reaches 0.811 with an inference time of less than 300 ms.

Conclusions: This study provides an accurate and fast drug detection solution for the management of special medications. The proposed rotated YOLO detector outperforms its YOLO counterpart in terms of precision.

Keywords: Toxic and narcotic medication; you only look once (YOLO); rotated object detection

Submitted Jun 21, 2022. Accepted for publication Jan 11, 2023. Published online Feb 24, 2023.
doi: 10.21037/qims-21-1146
View this article at: https://dx.doi.org/10.21037/qims-21-1146

Introduction

Medicinal toxic drugs and narcotic drugs especially, as well as psychotropic substances and radioactive pharmaceuticals, are special medications that require strict management in hospitals. If these medications inflow into the market illegally, the harm is unimaginable and includes the circulation of the crystal methamphetamine (1,2). In China, these four kinds of special medications are managed strictly by a special person, special counter, special account books, special prescription, and special book registration (3,4). As a daily task, these medications need to be double-checked under the supervision of two staff members every day. In parallel, the empty ampoules of these medications must be sent back to the central pharmacy after use to ensure that the quantity of prescribed drugs is consistent with that indicated by the empty ampoules. In the mode of traditional medicine cabinet management, many problems would arise. For instance, modification of anesthetic prescriptions is very common in clinics, which often leads to records that are inconsistent with the actual presence of medications in the counter because of the delay in the drug information update. Moreover, it is difficult to trace incidents such as false claims made by internal personnel, illegal prescriptions, and improper management of residual liquids (5,6). These problems bring great risks to the management of special medication, and there is an urgent need to implement informatization management of special medication with an intelligent medicine cabinet to reduce errors and avoid labor and time costs (7).

As a cornerstone project in computer vision with deep learning, object detection has long been attractive to researchers. The exploration and exploitation of these research achievements have stimulated the ongoing emergence of advanced products in the industry.

Ever since AlexNet won the ImageNet 2012 competition, deep learning has returned as a leading actor in the academic and industrial world (8). Spontaneously, researchers have proposed numerous algorithms to take advantage of deep learning in solving the object detection problem, which is a combination of classification and localization tasks. Ross Girshick designed a diagram leveraging features of the region-based convolutional neural network (R-CNN) (9). Endeavors of such a strategy include two stages. One is the extraction of deep features from input images, and the other is the determination of the region of interest (ROI) in which the extracted features contribute the most to the final prediction process. One intrinsic drawback of the R-CNN series is its relatively high latency, despite advancements in its siblings, such as fast-RCNNs (10) and faster-RCNNs (11). In 2015, Joseph Redmon invented the you only look once (YOLO) (12) style network, in which an input image is cut into different scaled grids, each containing multiple anchors in charge of the bounding box regression. In this way, the feature embedding and re-localization on ROIs could occur in one single round, which eliminates the need for the ROI searching process and tremendously decreases its time consumption. During the past few years, the YOLO series has continually evolved as a state-of-the-art object detector that not only wins titles in academic contests but also receives renown in the industrial field (13-16).

Most existing annotated data sets used for object detection projects provide bounding boxes as rectangles parallel to the x- and y-axes. However, in the real world, there are specific scenarios in which objects are usually densely arranged in arbitrary orientations. Under such circumstances arose special requirements, especially from the aerospace industry, in which aerial view photos are commonly used. Compared with the prediction of horizontal bounding boxes, rotated object detection has an extra variable-rotation angle θ to consider. Yang *et al.* provided solutions in which θ is encapsulated by specifically designed skew-intersection-over-union loss functions (17-20). The difference in the bounding box rotated angle created using this process contributes to the update of network parameters during the backpropagation procedure. Ming *et al.* made an effort to invent more efficient rotated anchor learning procedures and refined feature extractors that addressed the inconsistency between classification and bounding box regression and produce models that are more suitable for the arbitrarily oriented object detection task (21-23). Similar work related to aerial images was undertaken by Han *et al.*, who proposed the oriented detection module and feature alignment module (24,25).

Despite being a special case of rotated object detection, toxic and narcotic medication identification problems require thorough exploration. On the one hand, popular data sets used to train horizontal and rotated object detectors include some samples with the same distribution as the toxic and narcotic drugs of our interest. On the other hand, the existing rotated bounding box annotating method fails to provide the flexibility we hope to achieve. Moreover, the current non-maximum suppression (NMS) for rotated object detection is notorious for being a time-consuming post processor.

In this paper, we present an arbitrarily oriented object detector aimed at instantly and precisely identifying the specifications of toxic and narcotic medications. In this context, specification means a different medication or dose. The main contributions of this work boil down to three folds:

- YOLO-based backbone and head network have been revised and tuned to accomplish the toxic and narcotic medication detection target.
- A more flexible rotated bounding box annotation method is devised that shrinks the rotated angle into 0 to 90 degrees and forecasts the angle value employing a classification procedure.
- Last but not least, to make the proposed rotated toxic and narcotic medication detector feasible in

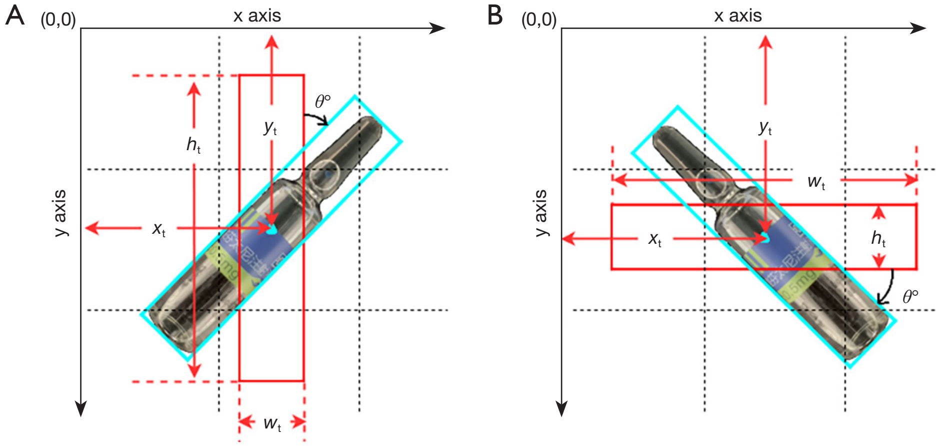

Figure 1 Illustration of rotated bounding box definition. The horizontal and rotated bounding boxes are marked in red and cyan respectively. The gray dashed lines indicate the grids in which multiple scaled and relocated anchors are given out as the predicted bounding boxes. The small value on the x-y-axis goes to the up-left corner. (A) and (B) show two kinds of horizontal bounding boxes that assures θ is less than 90°.

real application scenarios the rotated intersection over union (IOU) is calculated with help of a 0-1 mask for each found object.

This paper is organized as follows. A brief introduction of this work and the related articles are exhibited in Section 1. Section 2 explains the details of the proposed network. Extensive experiments and comparison results analysis go into Section 3. Finally, in section 4 we conclude this work.

Methods

Rotated bounding box

In previous rotated object detection researches in aerial imaging processing field, a bounding box with arbitrary rotated angle is defined based on the common scenes that the long edge lies along the x-axis and the rotated angle ranges from −90° to 90° (26,27). However, a rotation could have been described with degrees from 0° to 89°. Instead of following the traditional long-short edge prerequisite which would impose more memory consumption, in this work, we define the rotated bounding box more flexibly. In specific, the horizontal bounding box changes along when rotated angle is greater than 90°. As shown in *Figure 1*, the up-left corner is the zero point, $(x_t, y_t, w_t, h_t, \theta)$ designate a rotated bounding box on the image. x_t, y_t are coordinates of the box center point. w_t, h_t are width, and height on x- and y-axis respectively. (x_t, y_t, w_t, h_t) draws a horizontal rectangle that has nothing to do with θ. The black arrow points to the positive direction of the rotated angle. After the observation of 360° rotation it is concluded that 0° to 89° are enough to cover all the different situations as long as we change the width and height once the θ crossed the 90°. In *Figure 1*, the left and right parts indicate two layouts of an ampoule, both θ are 45°, though the two horizontal bounding boxes are in orthogonal places.

Just like the circular smooth label (26), the regression of a rotated bounding box is now transformed to be the regression of a horizontal bounding box and a rotated angle θ. Rather than predicting θ continuously, a classifier would be trained to decide which degree is the closest to θ. Different granularity could be taken when it comes to the assignment of rotation degrees. For example, 90 granularity ends up with 1° being represented by one class label, while 180 granularity uses 180 classes to identify 90 degrees which means each class represents 0.5°.

Toxic and narcotic medication detector

In this work, a YOLO-based architecture is leveraged to build the rotated toxic and narcotic medication detector. As shown in *Figure 2*, an input image fed into the network would go through the downward embedding and the upward re-localization processes in turn. The essence of YOLO is that multi-scale grids are imposed on the same image and each grid gives out 3 anticipations based on the predefined anchors. For training, not all the grids participate in the forward and back-propagation procedures.

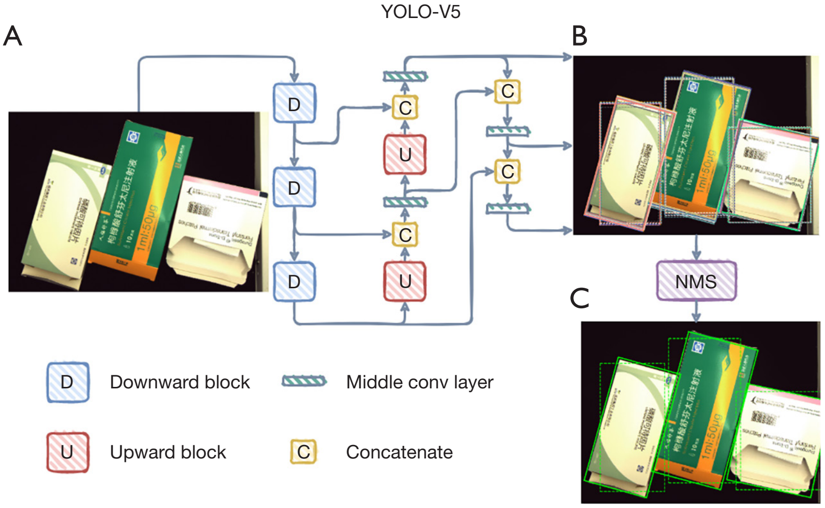

Figure 2 Schematic of the YOLO style toxic and narcotic medication detector. (A) shows the original input image. (B) is the output with all the predicted horizontal and rotated bounding boxes in dashed and solid lines respectively. The kept predictions after the NMS filter have been drawn in (C). YOLO, you only look once; NMS, non-maximum suppression.

Only those grids whose corresponding anchors are not too large or too small compared with the target bounding box are kept. The spaces covered by these grids are so called ROI. As a result, big sized objects would be taken care of by large scale grids and small sized objects fit in the denser grid scale. Moreover, when there are very few and small targets in the view, compared with the background, the number of ROIs would be limited. In case of the insufficiency of positive samples or the correctly predicted bounding boxes, two more neighbor grids of each target grid are picked up and assigned with the same label as the target grid's. The backbone of this proposed architecture follows the design of YOLO-V5 small edition (28).

The final output on each grid scale is a matrix $\boldsymbol{P} \in \mathbb{R}^{N_B \times N_A \times N_{G_h} \times N_{G_w} \times (5+N_C+N_D)}$, in which N_B is the number of batch size, N_A is the number of anchors for each grid, N_{G_h}, N_{G_w} are the number of grids on y and x-axis respectively, N_C is the number of medication specifications (classes), N_D is the number of degree intervals. Given H, W as the input image height and width, the predicted horizontal rectangle based on the jth anchor in the kth row, lth column grid of the ith batch index image could be depicted by the center point coordinates x_p, y_p, the width w_p and the height h_p of the bounding box. From Eqs. [1-7], $i \in [0, N_B)$, $j \in [0, N_A)$, $k \in [0, N_{G_h})$, $l \in [0, N_{G_w})$.

$$x_{\mathrm{p}} = lW/N_{G_w} + 2\sigma(P_{ijkl0}) - 0.5 \quad [1]$$

$$y_{\mathrm{p}} = kH/N_{G_h} + 2\sigma(P_{ijkl1}) - 0.5 \quad [2]$$

$$w_{\mathrm{p}} = 4\sigma(P_{ijkl2})^2 W_{\mathrm{A_j}} \quad [3]$$

$$h_{\mathrm{p}} = 4\sigma(P_{ijkl3})^2 H_{\mathrm{A_j}} \quad [4]$$

In Eqs. [1-4], σ refers to the sigmoid function $\sigma(\mathrm{x}) = 1/(1+\exp(-\mathrm{x}))$, $W_{\mathrm{A_j}}$, $H_{\mathrm{A_j}}$ are the width and height of the jth anchor, P_{ijkl0} means the first item in the last dimension of the predicted matrix, so as the P_{ijkl1} to P_{ijkl4}. Let IoU denote the IOU (29), the bounding box loss value for each grid is

$$L\_box_{ijkl} = \begin{cases} 1 - IoU\left((x_{\mathrm{p}}, y_{\mathrm{p}}, w_{\mathrm{p}}, h_{\mathrm{p}}), (x_{\mathrm{t}}, y_{\mathrm{t}}, w_{\mathrm{t}}, h_{\mathrm{t}})\right) & \text{if } x_{\mathrm{t}} \text{ existed} \\ \text{null} & \text{otherwise} \end{cases} \quad [5]$$

in which x_{t}, y_{t} are the coordinates of the target bounding box whose width is w_{t} and height is h_{t}. L_box is the mean value of all the $\mathrm{L\_box}_{\mathrm{ijkl}}$ who is not *null*.

Define $\boldsymbol{T} \in \mathbb{R}^{N_B \times N_A \times N_{G_h} \times N_{G_w}}$ which indicates whether or not there is an object in current grid.

$$T_{ijkl} = \begin{cases} 0 & \text{if } L\_box_{ijkl} \text{ is null} \\ 1 - L\_box_{ijkl} & \text{otherwise} \end{cases} \quad [6]$$

The objective loss value is the binary cross entropy between $\boldsymbol{P}$ and $\boldsymbol{T}$.

$$L\_obj = -\frac{\sum_{ijkl}\left[T_{ijkl}\log\left(\sigma\left(P_{ijkl4}\right)\right) + \left(1 - T_{ijkl}\right)\log\left(1 - \sigma\left(P_{ijkl4}\right)\right)\right]}{N_{\mathrm{B}} N_{\mathrm{A}} N_{\mathrm{G}_h} N_{\mathrm{G}_w}} \quad [7]$$

in which P_{ijkl4} indicates the objective loss is only affected by the fourth item in the last dimension of the predicted matrix. For each target label, there is a c_t representing the index of medication specification (class). Suppose there are n_t target boxes, let $\hat{\boldsymbol{T}} \in \mathbb{R}^{n_t \times N_C}$ be the one-hot label matrix.

$$\hat{T}_{ij} = \begin{Bmatrix} 1 & \text{if } j == c_t \\ 0 & \text{otherwise} \end{Bmatrix} \qquad [8]$$

in which $i \in [0, n_\mathrm{t})$, $j \in [0, N_\mathrm{C})$. The corresponding selected output could be represented as $\hat{\boldsymbol{P}} \in \mathbb{R}^{n_t \times N_C}$ each row of which are the 5th to ($\mathrm{N_C}$−1)th items in P. The classification loss value is

$$\mathrm{L\_cls} = -\frac{\sum_{ij}\left[\hat{T}_{ij}\log\left(\sigma\left(\tilde{P}_{ij}\right)\right)+\left(1-\tilde{T}_{ij}\right)\log\left(1-\sigma\left(\tilde{P}_{ij}\right)\right)\right]}{\mathrm{n_t N_C}} \qquad [9]$$

Similarly, to forecast the rotated degree θ, a one-hot matrix $\tilde{\boldsymbol{T}} \in \mathbb{R}^{n_t \times N_D}$ is introduced as

$$\tilde{T}_{ij} = \begin{Bmatrix} 1 & \text{if } j == \theta N_D/90 \\ 0 & \text{otherwise} \end{Bmatrix} \qquad [10]$$

in which $i \in [0, n_\mathrm{t})$, $j \in [0, N_\mathrm{D})$. Let $\tilde{\boldsymbol{P}} \in \mathbb{R}^{n_t \times N_D}$ be the corresponding selected rows with the $\mathrm{N_C}$th to ($\mathrm{N_D}$−1)th columns from $\boldsymbol{P}$. The loss value on degree is proposed as

$$\mathrm{L\_\theta} = -\frac{\sum_{ij}\left[\tilde{T}_{ij}\log\left(\sigma\left(\tilde{P}_{ij}\right)\right)+\left(1-\tilde{T}_{ij}\right)\log\left(1-\sigma\left(\tilde{P}_{ij}\right)\right)\right]}{\mathrm{n_t N_D}} \qquad [11]$$

The total loss value throughout the whole grid scales is a weighted summary of these four kinds of losses. Given $\mathrm{n_s}$ grid scales, let L_box(i), L_obj(i), L_cls(i), L_θ(i) represent the loss values in the *i*th space, the loss in all is

$$\begin{aligned}\mathrm{L} &= \gamma_{\mathrm{box}}\sum_{i=1}^{n_s} L\_box(i) + \gamma_{\mathrm{obj}}\sum_{i=1}^{n_s}\xi_i L\_obj(i) \\ &\quad + \gamma_{\mathrm{cls}}\sum_{i=1}^{n_s} L\_cls(i) + \gamma_{\theta}\sum_{i=1}^{n_s} L\_\theta(i)\end{aligned} \qquad [12]$$

in which ξ denotes the weight for object loss in different grid scale, γ is used to balance the four loss values.

Masked NMS

One of the most challenging tasks when designing an object detector is how to impose the NMS filter on the found boxes. There is no difference between horizontal and rotated bounding box NMS except for the *IoU* calculation (29). For horizontal *IoU* it is not complicated to figure out the rectangle areas, while for rotated *IoU* the intersection and union areas are irregular polygons which are not easy to get. Pure Cartesian geometry is a straightforward way to have the *IoU* result by calculating the polygon area of the intersection and the union, but it requires extremely high professional expertise to deal with the uncertainty of the shape. Moreover, such polygon area calculation could be time consuming. Alternatively, in this work, we devise an approximation leveraging 0-1 masks to find the *IoU* between two rotated bounding boxes. The idea is intrinsically straightforward and similar to former works like the mask-scoring (30) and the Pixels-IoU (PIoU) (31). *Table 1* illustrates the mechanism of our proposed solution. Given two boxes depicted by the quintet (x, y, w, h, θ), the mask could be received utilizing *get mask*, and the *iou ro* is in charge of getting *IoU* between these two rotated boxes. $\mathrm{h_m}$, $\mathrm{w_m}$ are hyperparameters deciding the mask size.

Results

Dataset and preprocessing

There are 38 toxic and narcotic medication specifications in Peking University Shenzhen Hospital. To alleviate the pressure from manual annotation, we randomly mix the single specification images to generate more annotated samples. The random manners include rotation, resize, perspective transformation, random background, random location, and repetition. Because the source images used for data generation are caught in unrestricted circumstances which may be distinct to the environment of our product, we separate the training process into two stages. In the first stage, 6,000 randomly generated images and their corresponding labels (also automatically generated) are put together with another 523 manually annotated photos. These data are fed into the model for training from scratch. Then the pre-trained model is fine-tuned on the same 523 manually annotated images again as the second stage. The first stage training comes by a non-overfitting network based on a relatively large dataset, and the second stage training is focusing on real-world dataset. Forty-two manually annotated photos are left for inference experiments. Each object has both horizontal and rotated bounding box labels.

All the images in the dataset are in RGB format with 0 to 255-pixel values. The shape of input data is 640×640. Each sample would go through the 255 division and the min-max normalization individually. When it comes to data augmentation, we exert a series of configurable maneuvers including mosaic transformation, moving, shrinking, rotation, cropping, horizontal and vertical flipping, augmentation on hue, saturation, and value. For bounding box augmentation, we do not leverage the rotation process while for rotated bounding box the shrink ratios on the x-

Table 1 *IoU* calculation for rotated bounding boxes

```
$\boldsymbol{get\_mask}(x, y, w, h, \theta)$:
    Get coordinates of up-left and bottom-right corners $x_{\min}, y_{\min}, x_{\max}, y_{\max}$ from $x, y, w, h$
    Let $\boldsymbol{M} \in \mathbb{N}^{h_m \times w_m}$ be the mask, $M_{ij} = 0, i \in [0, h_m), j \in [0, w_m]$
    if $\theta$ ==0:
    $M_{ij} = 1, i \in [y_{\min}, y_{\max}), j \in [x_{\min}, x_{\max})$
    else:
    Get slopes and intercepts of rectangle edges, say $a_0, b_{00}, b_{01}, a_1, b_{10}, b_{11}$
    $M_{ij} = 1, i, j \in (i > a_0 j + b_{00}) \& (i > a_0 j + b_{01}) \& (i > a_1 j + b_{10}) \& (i < a_1 j + b_{11})$
    return M
$\boldsymbol{iou\_ro}(x_0, y_0, w_0, h_0, \theta_0, x_1, y_1, w_1, h_1, \theta_1)$
    $\boldsymbol{M\_0} = \boldsymbol{get\_mask}(x_0, y_0, w_0, h_0, \theta_0), \boldsymbol{M\_0} \in \mathbb{N}^{h_m \times w_m}$
    $\boldsymbol{M\_1} = \boldsymbol{get\_mask}(x_1, y_1, w_1, h_1, \theta_1), \boldsymbol{M\_1} \in \mathbb{N}^{h_m \times w_m}$
    Get intersection $I = \sum_i^{h_m} \sum_j^{w_m} M\_0_{ij} M\_1_{ij}$
    Get union $U = \sum_i^{h_m} \sum_j^{w_m} M\_0_{ij} + \sum_i^{h_m} \sum_j^{w_m} M\_1_{ij}$
    $IoU = I/U$
    return IoU
```

IoU, intersection over union.

and y-axis are the same.

Configuration

This work has been developed under the environment of a single RTX 3060Ti GPU card and PyTorch framework. As mentioned above, the backbone architecture is following the YOLO-V5-S (28). Number of grid scales n_s =3, the respective grid steps are 8, 16, 32 on both axes. Correspondingly, the objective loss weight ξs are 4, 1, and 0.4. The balance parameters γ_{box} =0.05, γ_{cls} =0.5, γ_{obj} =1, γ_θ =0.5. Target box edges longer than 4 times or smaller than 1/4 times of anchor edge would be ignored, because they are too large or too small compared with the grid size. 180° and 90° angel granularities have been tested. For comparison, we have trained both models in horizontal and rotated versions to see the tradeoff between precision and efficiency on these two kinds of networks. There are 200 epochs training on the 6,000+523 dataset and another 200 epochs on the 523 datasets. The batch size is 1 for training and validation. Optimizer is Adam (32) starting with learning rate of 1e–3. The optimizer scheduler decreases the learning rate to 50% once no improvement was seen in the loss value of 10 epochs. For NMS, the confidence threshold is 0.45 for both horizontal and rotated scenarios, the *IoU* thresholds are 0.45 and 0.25 respectively.

Result analysis

In this section, comparisons have been undertaken on YOLO-V5 tailored for toxic and narcotic medication detection tasks and the proposed arbitrary oriented bounding box detectors with θ=90 and θ=180 respectively. Firstly, the influence of mask size h_m, w_m on the performance of models are unveiled.

Figure 3 exhibits the evaluation with the inference dataset on multiple mask sizes (h_m = w_m). Both inference speed in terms of frame per second (FPS) and accuracy in terms of mean average precision (mAP)[1] show a jump at the beginning and stays calm for the rest. This phenomenon is caused by the fact that small masks, despite coming with less memory reservation, would result in more ambiguous

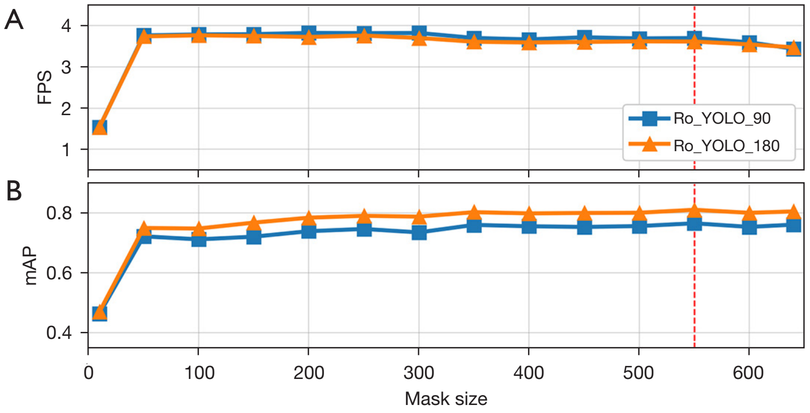

Figure 3 Comparison results with different mask sizes. (A) is for frame per second, (B) is for mean average precision. The two rotated degree modes are indicated as 'Ro_YOLO_90' and 'Ro_YOLO_180'. The red dashed line indicates the chosen mask size where $h_m = w_m = 550$. FPS, frame per second; mAP, mean average precision.

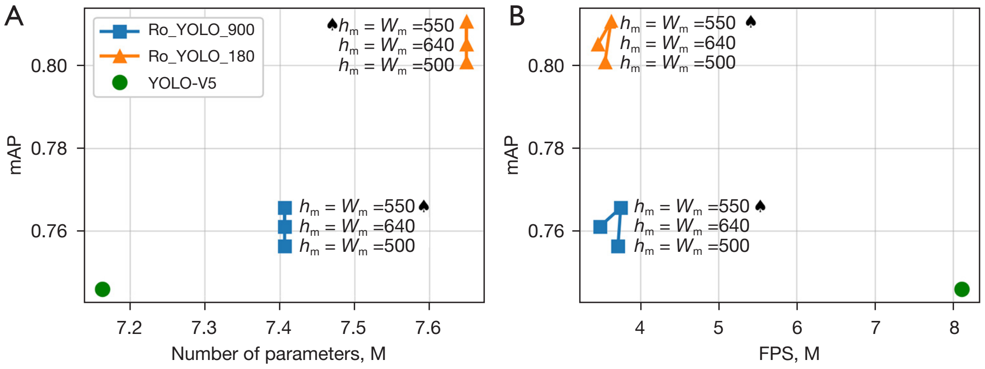

Figure 4 Performance comparison on different models. The y-axis is shared by two plots. Green circles represent tailored YOLO-V5, blue squares indicate rotated YOLO with θ=90, orange triangles are rotated YOLO with θ=180. Three mask sizes have been illustrated in this figure, and the black spade highlights the chosen configuration. (A) The mean average precision and number of parameters of each model; (B) the mean average precision and frame per second metric of each model. FPS, frame per second; mAP, mean average precision.

overlapped bounding boxes. Consequently, we choose $h_m = w_m = 550$ as default for the testing on two rotated YOLO models with different θs.

The tradeoff between accuracy and efficiency has been illustrated in *Figure 4*, in which the model size measured by the number of parameters, FPS, and mAP on different configured models are presented. The pros and cons of horizontal and arbitrary oriented object detectors are obvious to figure out. The best shot mAP of rotated YOLO model with θ=180 and $h_m = w_m = 550$ outperforms that of YOLO-V5 more than 5 percents. On the other hand, rotated detectors have hundreds of thousands of more parameters than their counterparts and need a longer time for inference.

In *Figure 5* we give out a few examples of the toxic and narcotic medications accompanied by the detected results. Although both the YOLO-V5 and rotated YOLO detectors could figure out most of the medication specifications

[1] Throughout this work, we follow the COCO dataset (33) tradition to calculate average precisions on different *IoU* thresholds ranging from 0.5 to 0.95 with 0.05 step. AP_{50}, AP_{75}, AP_{95} refer to the mean value of average precision for the whole classes when *IoU* is 0.5, 0.75, and 0.95 in respective. mAP is an average of all APs.

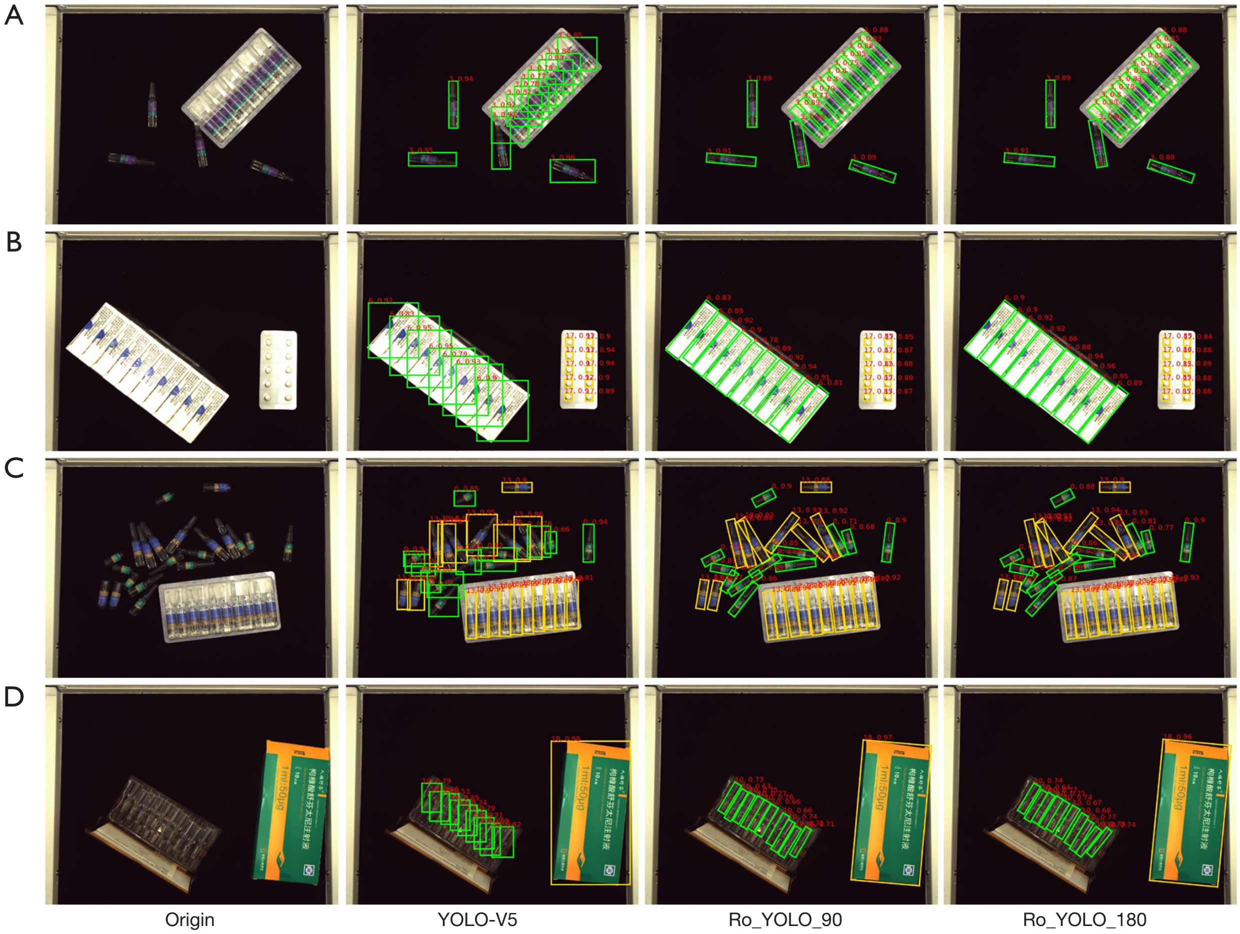

Figure 5 Detected results of different toxic and narcotic meditation detectors. The horizontal and rotated rectangles indicate the identified (brand-new and used) ampoules, pills, and boxes. In each photo, the same color represents one specification. The red labels mark the specification classes and the confidence ratio. (A) Ampoules and board packaged ampoules; (B) pills in board and box packaged; (C) different kinds of ampoules and board packaged ampoules; (D) different kinds of ampoules in board and box packaged.

correctly, it is easy to see the rotated YOLO detectors are better at identifying densely arranged drugs. In conditions of (A), (B), and (D), when the drugs in parallel are placed in a larger inclination angle, YOLO-V5 is hardly to precisely mark all the targets partly because there are large intersections existing among horizontal rectangles. Since most ampoules encapsulated injections are put in plastic trays, for instance, the sufentanil citrate injection (blue ampoule) and the pethidine hydrochloride injection (purple ampoule) as shown in *Figure 5*, the rotated YOLO detectors could provide more feasible predictions.

Table 2 exhibits more details about the contrast test, in which the rotated models are configured with $h_m = w_m = 550$. As the footnote in Section 3.3 explained, the average precision when IoU is 0.5, 0.75, and 0.95 are listed. mAP is the mean value all over the average precisions. Moreover, we also calculate the precision, recall, and F1 values. Without be distinguished as different objects, the whole classes of medications are taken into account to get a single value for each confidence and *IoU* threshold. Further, the mean value throughout these single values is kept as what has been presented in *Table 2*. The mAPs are in accordance with that of *Figure 4* which demonstrates the advantage of rotated detectors in terms of accuracy. When the *IoU* threshold increased, the average precision of rotated models, especially the one with θ=180, deteriorates slower than YOLO-V5. According to precision, recall, and F1 values we could find that rotated YOLO detectors outperform YOLO-V5 mostly in recall which indicates the rotated models could find more bounding boxes correctly.

Table 2 Contrast test results on different models.

Model	mAP	AP_{50}	AP_{75}	AP_{95}	Precision	Recall	F1
YOLO-V5	0.728	0.924	0.798	0.227	0.801	0.648	0.7
Ro_YOLO_90	0.766	0.97	0.817	0.222	0.796	0.656	0.693
Ro_YOLO_180	0.811	0.984	0.926	0.267	0.813	0.689	0.722

AP, average precision; mAP, mean average precision.

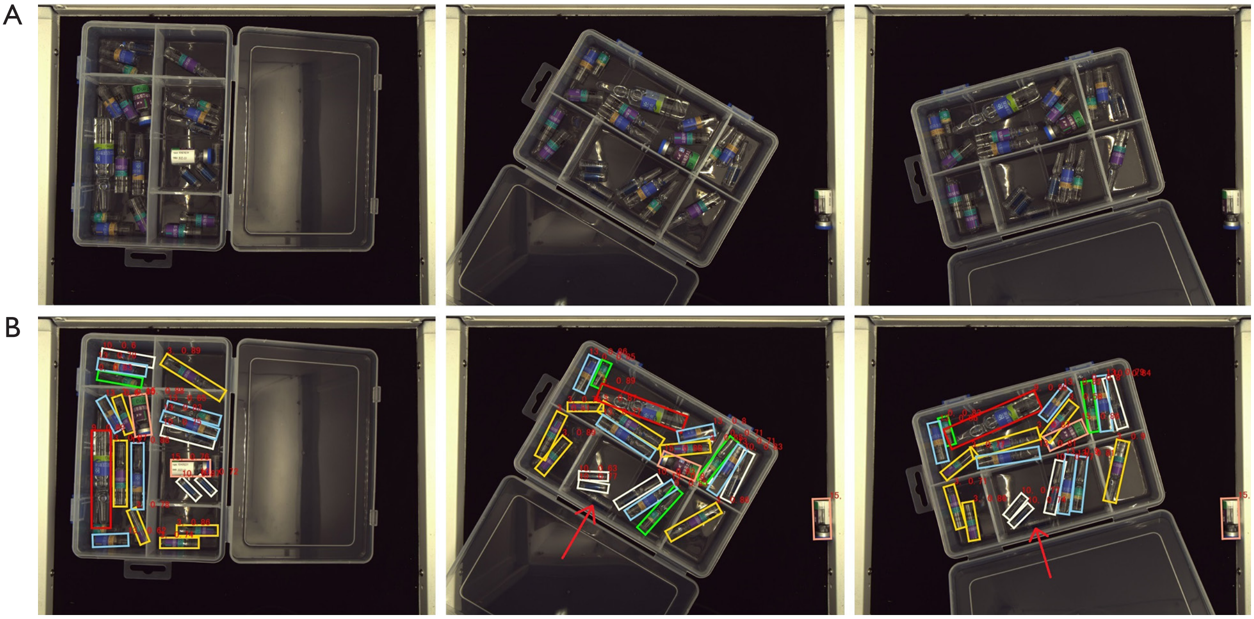

Figure 6 Detected toxic and narcotic drugs by the Ro_YOLO_180 detector in combo packaged box scenario. Colors indicates different drugs. (A) The original images; (B) Images with detected results. The red arrows indicate the omitted ampoule bottle.

The same evidence has been proven in *Figure 5*.

Discussion

From *Figure 5* we could tell another advantage of the rotated YOLO detectors that the used ampoules could be figured out by the length of the bounding box, which in contradictory is incapable for the horizontal bounding box detectors. Such feature is possible to be leveraged for further development of automatic devices. *Figure 6* illustrates one of the potential advancements that identifies both used and unused ampoules and vials in a combo packaged box. The rotated YOLO detector would predict the categories of the objects, which in this case are the drugs. According to the provided class-id-to-drug-name dictionary, we could give out what kind of drugs and how many of them are there in the menu box. The upper-level software could compare the result with the prescription in the database. Nurses could get eased from the time-consuming works of drug counting and checking and pay more attention on other more important works. In *Figure 6*, except for one morphine ampoule has been omitted, all the drugs have been correctly detected. Referring to the length of rotated bounding boxes it is easy to find the used bottles. More efforts could be spared to improve the precision of the detector in future works.

Recently, new trends of deep learning advancement in object detection have attracted academic and industrial attention. For example, the successfully cross-field proved transformer has been transformed to fit the object detection task (34). Hinton's group demonstrates the feasibility of solving the object detection as an image captioning problem (35).

Besides the object detection field, the prevalence of deep learning in other medical subjects also promotes the development of medical care (36,37). All these inspiring innovations shed light on the possibility for us to provide more effective and efficient products for the medical application community.

Conclusions

In this paper, the YOLO-based object detectors for toxic and narcotic meditation detection have been developed. With the more flexible rotated bounding box annotation method and the mask NMS at hand, we propose a medication detection network with arbitrarily oriented bounding boxes. Extensive contrast tests have demonstrated the feasibility and efficiency of the designed models.

Acknowledgments

Funding: This work was kindly funded by the National Natural Science Foundation of China (81660479), Bethune Charitable Foundation (B-19-H-20200622), Medical Science and Technology Research Foundation of Guangdong province, China (A2020157), and Peking University Shenzhen Hospital (JCYJ2018009).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-21-1146/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. No ethical approval or informed consent is required because of the nature of this study.

References

1. Hall W. The future of the international drug control system and national drug prohibitions. Addiction 2018;113:1210-23.
2. Rozenek EB, Wilczyńska K, Górska M, Waszkiewicz N. Designer drugs – still a threat? Przegl Epidemiol 2019;73:337-47.
3. State Council of the People's Republic of China. Regulations on the Control of Narcotic Drugs and Psychotropic Substances. 2005.
4. Chen S, Zhen C, Shi L. Study on the changing process of the narcotic drugs and psychotropic substances catalogues of China (1949-2019). Chinese Journal of New Drugs 2021;30:989-96.
5. Shuai L, Luo X, Ma X. Application Effect and Experience of Hospital Narcotic Drug Information Management System. Chinese Journal of Health Informatics and Management 2019;16:618-20.
6. Wang X, Yan D. Application of Multifunctional Narcotic Drug Management Cabinets in Operating Rooms. Pharmaceutical and Clinical Research 2018;26:318-20.
7. Wang Z, Ouyang P. Application and Practice of Poison and Hemp Based on Integrated Platform and Internet of Things. China Digital Medicine 2017;12:97-9.
8. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2012;60:84-90.
9. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2014:580-7.
10. Girshick R. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV); 2015:1440-8.
11. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137-49.
12. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2016:779-88.
13. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2017:6517-25.
14. Redmon J, Farhadi A. YOLOv3: An incremental

improvement. arXiv preprint, 2018. arXiv: 1804.02767.
15. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint, 2020. arXiv: 2004.10934.
16. Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: Exceeding YOLO Series in 2021. arXiv preprint, 2021. arXiv: 2107.08430.
17. Yang X, Yang J, Yan J, Zhang Y, Zhang T, Guo Z, Sun X, Fu K. SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019. doi: 10.1109/ICCV.2019.00832.
18. Yang X, Yan J, Feng Z, He T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object.arXiv preprint, 2019. arXiv: 1908.05612.
19. Qian W, Yang X, Peng S, Yan J, Guo Y. Learning Modulated Loss for Rotated Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35:2458-66.
20. Yang X, Yang X, Yang J, Ming Q, Wang W, Tian Q, Yan J. Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence. arXiv preprint, 2021. arXiv: 2106.01883.
21. Ming Q, Zhou Z, Miao L, Zhang H, Li L. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35:2355-63.
22. Ming Q, Miao L, Zhou Z, Dong Y. CFC-Net: A Critical Feature Capturing Network for Arbitrary Oriented Object Detection in Remote-Sensing Images. IEEE Transactions on Geoscience and Remote Sensing. 2021:1-14.
23. Ming Q, Miao L, Zhou Z, Song J, Yang X. Sparse Label Assignment for Oriented Object Detection in Aerial Images. Remote Sens 2021;13:2664.
24. Han J, Ding J, Li J, Xia GS. Align Deep Features for Oriented Object Detection. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1-11.
25. Han J, Ding J, Xue N, Xia GS. ReDet: A Rotation-Equivariant Detector for Aerial Object Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021:2786-95.
26. Yang X, Yan J. Arbitrary-Oriented Object Detection with Circular Smooth Label. In: Proceedings of the European Conference on Computer Vision (ECCV); 2020:677-94.
27. Yi J, Wu P, Liu B, Huang Q, Qu H, Metaxas D. Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2021:2150-9.
28. Glenn J. YOLO-V5. 2021. Available online: https://github.com/ultralytics/yolov5
29. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34:12993-3000.
30. Huang Z, Huang L, Gong Y, Huang C, Wang X. Mask Scoring R-CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019:6402-11.
31. Chen Z, Chen K, Lin W, See J, Yu H, Ke Y, Yang C. PIoU Loss: Towards Accurate Oriented Object Detection in Complex Environments. In: Proceedings of the European Conference on Computer Vision (ECCV); 2020:195-211.
32. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint, 2014. arXiv: 1412.6980.
33. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common Objects in Context. In: European Conference on Computer Vision – ECCV 2014. Cham: Springer International Publishing; 2014:740-55.
34. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint, 2021. arXiv: 2103.14030.
35. Chen T, Saxena S, Li L, Fleet DJ, Hinton G. Pix2seq: A Language Modeling Framework for Object Detection. arXiv preprint, 2021. arXiv: 2109.10852.
36. Ardhianto P, Subiakto RBR, Lin CY, Jan YK, Liau BY, Tsai JY, Akbari VBH, Lung CW. A Deep Learning Method for Foot Progression Angle Detection in Plantar Pressure Images. Sensors (Basel) 2022;22:2786.
37. Tsai JY, Hung IY, Guo YL, Jan YK, Lin CY, Shih TT, Chen BB, Lung CW. Lumbar Disc Herniation Automatic Detection in Magnetic Resonance Imaging Based on Deep Learning. Front Bioeng Biotechnol 2021;9:708137.

Cite this article as: Peng J, Wang F, Ma X, Chen Z, Fu Z, Hu Y, Zhou X, Wang L. Towards toxic and narcotic medication detection with rotated object detectors. Quant Imaging Med Surg 2023;13(4):2156-2166. doi: 10.21037/qims-21-1146