

# Pill Detection with Medical Drug Identification System

*Prajwal S*  
Computer Science and Engineering  
Alva's Institute Of Engineering  
Mangalore, India  
prajwalkkp03@gmail.com

*Nikhil Lobo*  
Computer Science and Engineering  
Alva's Institute Of Engineering  
Mangalore, India  
nikhillobo1609@gmail.com

*Rushikesh R Yaligar*  
Computer Science and Engineering  
Alvas's Institute Of Engineering  
Mangalore, India  
rushikeshry0202@gmail.com

*Navaneeth Surya P*  
Computer Science and Engineering  
Alva's Institute Of Engineering  
Mangalore, India  
nsurya5302@gmail.com

**Abstract** Identifying drug pills is one of the most crucial duties for medication safety. Improving pharmaceutical safety starts with accurately identifying drugs based on their outward appearance. The goal of earlier research was to identify drugs by looking at them from either the front or the back at a set viewing angle. In actual applications, the prior approaches have trouble identifying and detecting distinct substances in circumstances when there are several medications and the drugs are put randomly. In this work, a convolution neural network-based detector is developed to overcome the challenges and help patients identify drugs. A localization stage and a classification stage are part of the suggested system. For drug localization, the enhanced feature pyramid network (EFPN) is suggested, while for drug classification, Inception-ResNet v2 is employed. Sixty-one medication dataset categories are included in the proposed medication Pills Image Database, which is intended to support deep learning studies in the pharmaceutical industry. In the localization experiment, the suggested EFPN attains an accuracy rate of more than 96%. The suggested system received the Top-1, Top-3, and Top-5 accuracies of 82.1, 92.4, and 94.7%, respectively, in the entire system evaluation.

## I. INTRODUCTION

Six to eight thousand lives have been lost annually due to medication errors [[1]]. Patients frequently struggle to tell the difference between unpackaged medications and self-medication errors. Furthermore, statistical studies have calculated that 3–7% of prescribed drugs are not utilized, resulting in an annual cost to the US of about \$5 billion [[2], [3]]. The elimination of medicine waste has made improving medication knowledge and giving patients sufficient pharmacological information crucial [[4]]. Nonetheless, patients still find it challenging to identify drugs based only on their appearances.

The primary obstacles in drug identification stem from the vast array of variations and striking similarities among the medicines. In a medical center, for instance,

there can be up to 700 different drug categories, many of which would look the same to an untrained eye. In general, the form, color, and imprint of a substance can influence how it appears. Several varieties of drug forms can be distinguished, including round, capsule, oval, barrel, 4-sided, 6-sided, and so forth. Regarding the medication colors and impressions, additional distinguishing characteristics are also discernible. In traditional drug identification, the user must spend a lot of time manually inputting the drug's characteristics on the drug website in an attempt to identify an unknown substance. Drug recognition has not been able to adequately address a number of issues in earlier research [[5]-[9]], including the occurrence of several drugs in one image and random drug placements. Determining and normalizing the drug rotation angle for every drug category presents a challenge as well. Moreover, automatic pill segmentation using conventional image processing is challenging.

The drug detection system in our earlier study [[10]] uses Xception [[12]] and Feature Pyramids Networks (FPNs) [[11]] for drug categorization and localization. In this paper, we offer Enhanced Feature Pyramid Networks (EFPNs), which are used to improve drug localization accuracy by combining the FPN with Global Convolution Network (GCN) [[13]]. Additionally, the drug image database is expanded and the Inception-ResNet v2 is replaced with Xception in the current effort for drug classification. The Kaohsiung Veterans General Hospital (KVGH) provides the photographs used to create the Drug Pills Image Database (DPID). The database includes training, validation, and testing datasets and has 2,429,753 images altogether across 612 pharmacological categories.

The following is a summary of this work's primary contributions: (i) The issue of multiple drug localization is suggested to be resolved by the EFPN. (ii) The multi-category classification problem with the variables, such as the fluctuations in light angles and drug rotations, is solved by a convolution neural network (CNN)-based classifier. (iii) The drug database DPID, which is collocated for deep learning, has pictures with different characteristics, like different light angles and pill rotations, as well as many tablets in a single image. This is how the paper is structured. Section 2 discusses earlier research on CNN-based object detectors, drug databases, and drug pill detection.

## II. LITERATURE SURVEY

In recent years, significant strides have been made in medical image segmentation and medication-related technologies. Chiun-Li Chin et al. proposed a novel fuzzy DBNet for medical image segmentation, offering promising advancements in this critical area. Their work, published in *Electronics*, introduces an innovative approach aimed at enhancing the accuracy of medical image segmentation, thereby potentially improving diagnostic processes and treatment planning.

Meanwhile, Khalil Al-Hussaeni, Ioannis Karamitsos, and their team focused on CNN-based pill image recognition for retrieval systems, as detailed in their study published in *Applied Sciences*. Their research addresses the growing need for efficient pill identification methods, particularly in retrieval systems, which play a vital role in healthcare management and medication adherence.

In a different domain, Alexandru G. Berciu, Eva H. Dulf, and Iulia A. Stefan introduced a flexible augmented reality-based health solution for medication weight establishment, showcased in *Processes*. This innovative solution holds promise for improving medication dosing accuracy, potentially enhancing patient safety and treatment effectiveness.

Moreover, Tony Salloom, Okyay Kaynak, Xinbo Yu, and Wei He explored the application of a proportional integral derivative booster for neural networks-based time-series prediction, focusing specifically on water demand prediction. Their work, published in *Engineering Applications of Artificial Intelligence*, underscores the significance of accurate prediction models in various domains, including resource management and infrastructure planning.

In the realm of pharmaceutical technology, Xuesan Su, Yaonan Wang, Jianxu Mao, and their colleagues conducted a review of pharmaceutical robots based on hyperspectral technology, contributing insights into the potential applications and advancements in this burgeoning field. Their work, published in the *Journal of Intelligent & Robotic Systems*, sheds light on the evolving landscape of pharmaceutical automation and its impact on healthcare delivery.

Furthermore, KyeongMin Cha, Hyun-Ki Woo, Dohyun Park, and Dong Kyung Chang delved into the effects of background colors, flashes, and exposure values on the accuracy of a smartphone-based pill recognition system. Their study, published in *JMIR Medical Informatics*, underscores the importance of optimizing image capture conditions for enhancing the performance of pill recognition technologies, ultimately aiming to improve medication management and adherence.

Finally, Urja Patel presented findings from the 2021 International Conference on Artificial Intelligence and Smart Systems, offering insights into the latest advancements and trends in AI and smart systems. Additionally, Alexandru G. Berciu, Eva H. Dulf, and Iulia A. Stefan introduced PillCrop, a solution designed to facilitate the correct administration of medicine, as outlined in *IFAC-PapersOnLine*.

## III. RELATED WORK

This section provides an overview of drug pill detection, drug databases, and CNN-based object detectors.

### A. Automatic drug pills detection and drug database

A few investigations on pill acknowledgment frameworks have been proposed for the order of medications. In 2010, Hartl [[14]] involved the shape and variety boundaries as question highlights for an obscure medication. Lee et al. [[15]] considered the Scale-Invariant Element Change and Multi-Scale Nearby Twofold Examples for drug coordinating. In 2012, Caban et al. [[16]] changed the shape dispersion to look at the shape, variety and engraving of medications and made an invariant descriptor for drug acknowledgment. The pill's engraving was portrayed by the weighted shape setting in [[17]]. Chen et al. [[9]] proposed a comparability estimation for the shape and shade of medications, where the shape order was applied utilizing a basic brain organization, and the shades of the medications were moved from the RGB variety space to the HSV variety space. Yu et al. [[8], [18]] proposed a changed stroke width change to identify hints of the impression for engrave highlights. Suntronsuk and Ratanotayanon [[19], [20]] proposed a K-implies based grouping to extricate the engraving text for drug order. Neto et al. [[5]] proposed an invariant component extractor in view of the shape and shade of the medications. As indicated by the above scholarly works, the shape, variety, and engraving of a medication are the critical qualities for highlight extraction. Past examination works have all centered around single medication arrangement in view of the elements at the outdoors of the pill, and the pill districts inside a picture have been resolved physically or by restricting the foundation tone to dark [[5]-[9]].

Zeng et al. [[6]] proposed the MoblieDeepPill, which utilized the Histogram of Arranged Angle and Backing Vector Machines (SVMs) for single medication localisation and the multi-CNN models were applied to gather pill qualities. Wang et al. [[7]] utilized three GoogLeNet Origin models with various consequences for the variety, shape, and element, and choice combination was utilized to join those models. In any case, in [[7]], the places of pill pictures were restricted to turns by 90°, 180°, and 270°. In the commonsense utilizations of a medication pill discovery framework, the previously mentioned examinations would have issues requiring further turn of events, like the assurance of haphazardly situated and pivoted pills, as well as the quantity of pills in every recognition task.

The medication pictures utilized in many examinations have been gathered from different sites. A few scholarly works [[15]-[17], [20]] utilized recommended drug pictures from online pill data sets or sites, for example, 'drug.com', 'pharmer.org', the U.S Medication

Implementation Organization Office of Legal Sciences, the U.S. Public Library of Medication, etc. A few investigations developed pill datasets for the tests [[8], [9], [18]]. In [[9]], 526 medication pictures in 263 classifications were obtained. Yu et al. [[8], [18]] gathered 12,500 medication pictures in 2500 classes. Research works in [[5]-[7]] utilized the Public Library of Medication Picture Data set, NIH NLM PIR [[21]], which is a public data set containing front and back perspectives on 1000 medications. The data set contains 2000 top notch pictures and 3000 customer pictures. Albeit a few medication data sets have been gathered in past works, the general inclusion of the data sets is deficient for profound learning, particularly with respect to the progressions in lighting, revolution and medication plots for the different pill classes. The outline of the medication datasets is displayed in Table 1.

| Database                 | Reference         | Classes    | Training         | Validation    | Testing      |
|--------------------------|-------------------|------------|------------------|---------------|--------------|
| NIH NLM PIR              | [[5]-[7], [21]]   | 1000       | 2000             | 3000          | —            |
| captured by themselves   | [[8], [18]]       | 2500       | 12,500           | —             | —            |
| captured by themselves   | [[9]]             | 263        | 526              | —             | —            |
| captured by themselves   | [[10]]            | 131        | 488,520          | 8280          | a1680        |
| collected form website   | [[15]]            | —          | 2116             | —             | —            |
| collected form website   | [[16]]            | 568        | 568              | —             | —            |
| collected form website   | [[17]]            | 2000       | 2500             | >12,000       | —            |
| pill database 'drug.com' | [[19], [20]]      | —          | 540              | —             | —            |
| <b>DPID</b>              | <b>this study</b> | <b>612</b> | <b>2,393,585</b> | <b>34,975</b> | <b>a1193</b> |

### B. CNN-based object indicators

The area CNN (R-CNN) is an illustration of average item indicators [[22]-[26]]. The R-CNN technique utilizes particular pursuit to separate district proposition and the direct SVM is utilized for characterizations [[22]]. Quicker R-CNN utilizes the District Proposition Organization as opposed to the particular inquiry and uses anchor boxes to take care of the scale-variation issue [[25]]. The return for capital invested shrewd Opposite Reweighting Organization [[27]] is based upon the Quicker RCNN and comprises of a multi-facet pooling activity and return for capital invested insightful converse reweighting for various kinds of stamping. Cheng et al. joined the extra multi-point secures with the RPN to address the rotational varieties [[28]]. The FPN [[11]] was added to the spine for multiscale localisation in the Cover R-CNN [[26]]. RefineDet [[29]] incorporated the anchor refinement and the item location modules, and the exchange association block was proposed to build the element map for object discovery. Yuan et al. [[30]] proposed the Upward Spatial Arrangement Consideration (VSSA) Organization for traffic sign recognition, wherein the multi-goal highlight learning module and the VSSA module were given to improve the semantic of little size objects, and to acquire setting data on the article. The revolution invariant layer and a Fisher discriminative layer were added to the current item identification CNN to tackle the issues of pivot varieties, inside class variety, and between-class closeness in [[31], [32]].

Visual Math Gathering Net (VGG-16/VGG-19) [[33]] and profound remaining organization (ResNet-50/ResNet-101) [[34]] are the normal spines for object identification. The two organizations have won in arrangement undertakings of PC vision. In any case, the VGG and the ResNet can't act as classifiers for drug order, yet the ResNet is as yet a decent FPN spine for drugs localization

## IV. PROPOSED FRAMEWORK

The framework engineering of programmed drug pill location is displayed in Fig. 1. A two-stage engineering is taken on in the proposed framework, which incorporates the localisation and grouping stages. The medication not entirely settled through the EFPN in the localisation stage. The medications are delegated the closer view in the medication localisation stage, no matter what the shape, variety, and surface of the medications. Both the directions and sizes of the medications are assessed in the primary stage. The proposed EFPN, which gives the localisation in various scales, is built for object areas assessment. Two completely convolutional models, relapse and order, are taken on to decide the directions and sizes of the medications for object expectation. Non-greatest concealment (NMS) is applied to kill excess jumping boxes.

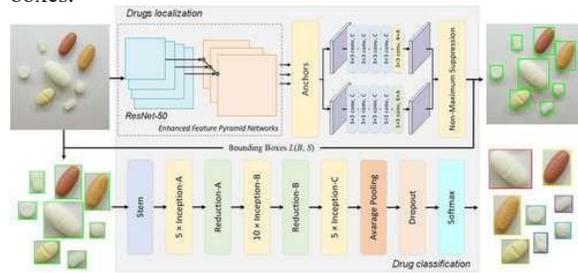


Fig. 1

System overview of automatic drug pills detection. The proposed system contains a two-stage architecture, localisation, and classification. In the localisation step, the proposed EFPN and ResNet-50 are used for backbone building; two sub-models are applied for bounding box estimation, non-maximum suppression is utilised to merge the prediction results and reduce bounding boxes. Inception-ResNet V2 is used for drug classification

Framework outline of programmed drug pills recognition. The proposed framework contains a two-stage design, localisation, and grouping. In the localisation step, the proposed EFPN and ResNet-50 are utilized for spine building; two sub-models are applied for jumping box assessment, non-most extreme concealment is used to blend the expectation results and decrease bouncing boxes. Origin ResNet V2 is utilized for drug order

Albeit the EFPN localisation model can distinguish the medication areas, the medication closeness pivot actually cause hardships in the EFPN grouping model. The rotational point is hard to characterize. Moreover, the surface and engraved text are conflicting between various medications. The CNN-based classifier successfully processes the surface subtleties on the item's

surface as per the preparation information and information expansion. Subsequently, the Commencement ResNet v2 is utilized for drug grouping all things being equal. The subtleties are examined in the accompanying.

#### A. Improved include pyramid network for object localization

The proposed EFPN has been constructed in view of the ResNet-50 spine. Five convolutional layers, C1, C2, C3, C4, and C5, have been remembered for the ResNet-50. Three of the convolution layers, C3, C4, and C5 have been picked as the benchmark for the pyramid layers. The layers of the ResNet give the progressive elements, yet the shallow layers are frail in giving the component portrayal. Subsequently, Lin et al. [[11]] have proposed the FPN for consolidating the semantics with the hierarchal elements.

As per the trial results on drug localisation, objects with bigger proportions are frequently imperceptible. The FPN is lacking for drug localisation. The EFPN produces two layers by the GCN [[13]] to expand the open field of FPN. The subtleties of the EFPN engineering are displayed in Fig. 2. The base up pathway, horizontal associations, hierarchical pathway, and the GCN have been remembered for the proposed EFPN design. The base up pathway is created by the forward spread of the ResNet-50 spine. The spine network decreases the size of element maps through convolution layers with step and pooling. The parallel association is made out of  $1 \times 1$  convolution layer with 256 channels, which are utilized to lessen the channel number and execute data combination across the channels. The hierarchical pathway delivers the component maps with more grounded semantics on the lower pyramid level, and the closest neighbors introduction has been used to up-example the hierarchical pathway. In the EFPN, the GCN has been utilized to assemble the pyramid layers P6 and P7 in view of the C5 layer. The GCN expands the open field of the FPN for the identification of bigger articles. The GCN factorizes the  $k \times k$  convolutions into  $k \times 1$  and  $1 \times k$  without the Redressed Direct Unit (ReLU) capability. The age of the pyramid layers P6 and P7 is displayed in Fig. 3. The pyramid layers P3, P4, P5, P6, and P7 have been built by the EFPN with the ResNet-50 as the spine. The pyramid layers are considered for object localisation with various scales and proportions. What's more, the pyramid layer P2 in the EFPN has been eliminated in light of the fact that the pyramid layers from P3 to P7 can cover the localisation task for conflicting medication sizes.

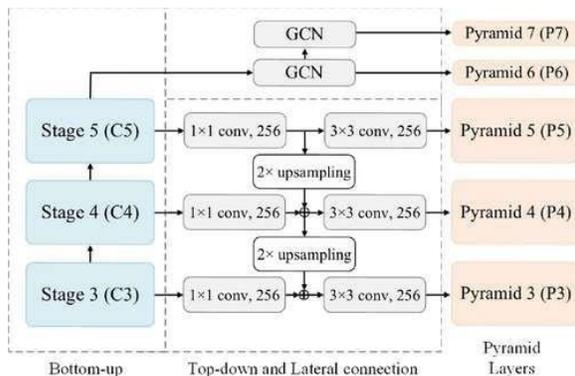


Fig. 2 Architecture of EFPN for the pyramid layers generation

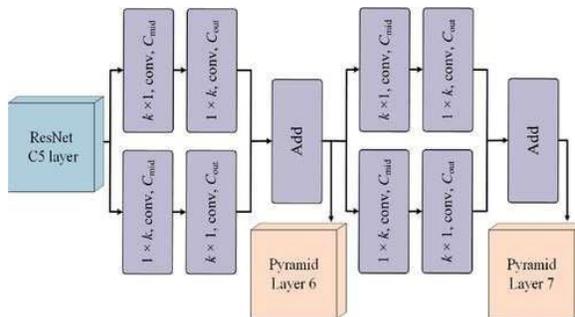


Fig. 3 Generation of pyramid layers P6 and P7

The anchors have been used to create the pre-chosen boxes for relapse and grouping models in every one of the pyramid layers. The anchors have been created to build the anchor thickness, through the numerous perspective proportions  $\{1:2, 1:1, 2:1\}$  and sizes  $\{20, 21, 22\}$ . The motivation behind the anchors is to cover different appearances of the bouncing boxes in each place of the component map. In the proposed framework, nine anchors have been made with various angle proportions and sizes at each direction point. The Crossing point over-Association (IoU) proportion is utilized to channel the abundance secures, which analyzes the likeness between the anchor and the close to ground-truth bouncing boxes. The IoU recipe is displayed in the accompanying condition:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

Consequently, a positive mark is relegated to an anchor in the event that the IoU proportion surpasses 0.5 when contrasted with the ground-truth box and a negative name is given on the off chance that the IoU proportion is under 0.4; generally the anchor is deserted. Both the relapse and the order of the bouncing box are determined when the anchor has a positive name. To prohibit the unessential anchors in the proposed framework, the quest regions for the IoU estimation are set to 322, 642, 1282, 2562, and 5122 on the pyramid layers P3, P4, P5, P6, and P7, separately.

To recognize the item area on the pyramid levels, a grouping model and a relapse model have been used. The reason for the grouping model is to anticipate the article classification of the anchors. The arrangement

model contains five CNN layers. The initial four layers are made out of  $3 \times 3$  convolution layers with C channels and ReLU initiation. The  $3 \times 3$  convolution layers with  $K \times A$  channels and straight initiation are utilized as the result layers. The boundaries K, C, and A, are the one-hot vector of the classifications, the quantity of channels, and the quantity of anchors, separately. The relapse model is utilized to appraise the bouncing box  $R_{ba}\{b_x, b_y, b_w, b_h\}$  as per the anchor  $R_{ra}$  and the ground-truth object  $R_{ga}\{g_x, g_y, g_w, g_h\}$ . The relapse model is like the order model, the main contrast is that the  $3 \times 3$  convolution layers with direct enactment and  $4 \times A$  channels are considered as the result layer. Concerning the bouncing box, the boundaries x, y, w, and h, individually, address the x and y facilitates, the width, and the level. The anchor is addressed as a, where  $a = 1, 2, \dots, A$ . The relapse model has been embraced to track down a planning  $f(R_a) = R_{ba} \simeq R_{ga}$  and the parameterisation of bouncing box [[22]] is taken as the reference in our relapse model. In each anchor  $R_{ra}$ , the anticipated jumping box  $R_{ba}\{b_x, b_y, b_w, b_h\}$  is determined by the general offset t and the anchor box  $R_{ra}$ . The anticipated jumping box  $R_{ra}$  is signified as follows:

$$\begin{aligned} b_x &= r_w t_x + r_{x_2} \\ b_y &= r_h t_y + r_{y_2} \\ b_w &= r_w \exp(t_w) \\ b_h &= r_h \exp(t_h) \end{aligned} \quad (10)$$

where the relative offset  $t\{t_x, t_y, t_w, t_h\}$  is estimated by the regression targets of the nearby ground-truth bounding box  $R_a\{n_x, n_y, n_w, n_h\}$  and the anchor box  $R_r\{r_x, r_y, r_w, r_h\}$ .

The regression target of the ground-truth bounding box  $R_a$  is estimated by the anchor  $R_r$  and the ground-truth object  $R^a = \{g_x, g_y, g_w, g_h\}$ , and defined as follows:

$$\begin{aligned} n_x &= (g_x - r_x) / r_w \\ n_y &= (g_y - r_y) / r_h \\ n_w &= \log(g_w / r_w) \\ n_h &= \log(g_h / r_h) \end{aligned} \quad (11)$$

To blend the anticipated aftereffects of the pyramid layers, the non-most extreme concealment (NMS) [[35]] is utilized to dispense with the cross-over IoU. The expectation results  $L(B, S) = \{b_n, s_n\}$  are arranged in diving request S, where n is the record of the anchor, with  $n = 1, 2, \dots, N$  and N is the quantity of the bouncing boxes. NMS handling has two stages. In the initial step, the jumping box which has the greatest certainty score M (bm, sm) is chosen and moved to the localisations list, D. In the subsequent step, the IoU is determined between the score, M (bm, sm) and each bouncing box of rundown L(B, S); and the jumping boxes are barred when the IoU is more noteworthy than the IoU edge TIoU. At last, the forecast results L(B, S) are assessed by various cycles of the two stages.

The localisation network is built by the relapse and order models, which are utilized to anticipate the position  $R_b\{b_x, b_y, b_w, b_h\}$  and the article classifications  $P = \{p_0, p_1, \dots, p_{K-1}\}$ . Both of the models are prepared by a perform various tasks misfortune capability  $LML(u, t, R_b, P, Y_{gt})$ , which is meant in the accompanying condition:

$$L_{ML}(u, t, R_b, P, Y_{gt}) = uL_{loc}(t, R_b) + L_{cls}(P, Y_{gt}) \quad (10)$$

where  $L_{loc}$  and  $L_{cls}$  are the relapse misfortune and the grouping misfortune, the boundary u mirrors the positive (target) or the negative (foundation) marking.  $R_n$  is the place of the close to ground-truth object target. The one-hot vector  $Y_{gt}$  is the portrayal of the ground-truth class mark. The relapse model takes the standard L1-standard misfortune for the bouncing box relapse. The equation is signified as follows:

$$\begin{aligned} L_{loc}(t, R_b) &= \sum_{i \in \{x, y, w, h\}} \text{Smooth}_{L1}(t_i - b_i) \\ \text{Smooth}_{L1}(x) &= \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \end{aligned} \quad (12)$$

An outrageous irregularity in the preparation stage might result since there are in every case more foundation tests than the forefront targets. The central misfortune approach [[36]] is utilized to tackle the awkwardness issue through the down-weighting inlier. The central misfortune approach comprises of a better cross-entropy CE (p, y) equation as displayed in (13)

$$\begin{aligned} CE(p, y) &= CE(p_t) = -\log(p_t) \\ p_t &= \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{if } y = 0 \end{cases} \end{aligned} \quad (13)$$

where y indicates the ground-truth class, and p is the assessed likelihood for the calls with name  $y = 1$ . This approach takes on the profound brain organization to zero in on the examples that are hard to anticipate, and the central misfortune is summed up as per the usually utilized cross-entropy by diminishing the punishments from the very much ordered examples. The central misfortune approach adds a regulating factor  $(1 - pt)\gamma$  to the cross-entropy with a customizable boundary  $\gamma$ . The condition for the central misfortune approach is displayed underneath:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (14)$$

The focal loss samples with small losses can dominate the gradient when the accumulated classification loss exceeds a large number of inliers. The classification loss  $L_{cls}$  is defined by the sum of the focal loss samples over K classes, and the formula  $L_{cls}$  is shown in the following equation:

$$L_{cls}(P, Y) = \sum_{i=1}^K FL(p_i, y_i)_{(16)}$$

where  $i$  is the index of the category,  $p_i$  and  $y_i$  are the ground-truth class and the probability, respectively.

### B. Origin ResNet v2 for drug characterization

The proposed framework endeavors to consolidate another CNN model for drug characterization. To further develop exactness, the classifier needs to serious areas of strength for have figuring out abilities. GoogLeNet [[12], [37], [38]] has gotten critical consideration in the grouping of an enormous number of classes by stacking origin modules and expanding the beginning modules. Beginning V3 [[37]], Xception [[12]], Initiation V4 [[38]], Origin ResNet v1 [[38]], and Commencement ResNet v2 [[38]] have all been assessed concerning drug characterization.

Origin ResNet v2 has been chosen for the medication characterization task due to its trial execution. Stem block is utilized as the spine, and three modules, named Origin A, Origin B, and Beginning C, have been utilized for highlight map combination in various scales. The engineering of Beginning ResNet v2 is depicted in the medication grouping period of Fig. 1. The stem and the three modules, Origin A, Beginning B, and Origin C are displayed in Figs. 4a and b.

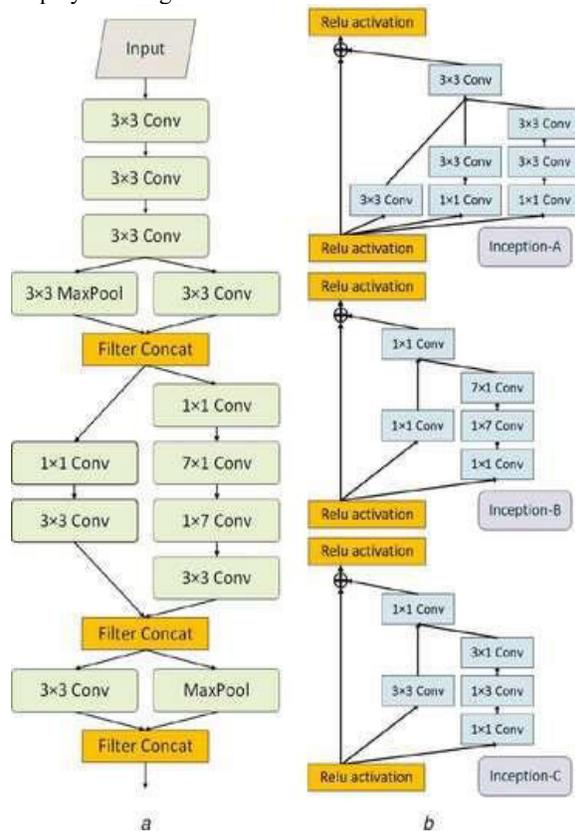


Fig. 4 Modules of Inception-ResNet v2

(a) Architecture of stem, (b) Those models, Inception-A, Inception-B, Inception-C, are represented from top to down

## V. EXPERIMENTS AND DISCUSSION

The proposed programmed drug pills recognition framework is assessed on the DPID. The medication pills location has a few difficulties, including various classifications, conflicting sizes, comparative appearances, and irregular revolution. In this part, we portray the data set and trial setting, as well as the unwavering quality assessment of the proposed framework, trailed when intricacy examination and a conversation of mixed up cases.

### A. Medication Pills Picture Information base

The DPID is gathered through the participation with the KVGH, which has been introduced basically to help research in PC vision of drug applications. To keep up with the picture quality and catch the engraving in the DPID, a computerized single-focal point reflex camera, Group EOS 80D, has been utilized for information assortment. Pictures with varieties in the light points, turn, and medication points have been considered in the DPID. 30-40 medication pictures have been caught from various plots for every class and the medication pictures are produced by turning the pictures from  $0^\circ$  to  $359^\circ$ .

The DPID contains 612 classifications of medications from the KVGH, and three datasets have been autonomously gathered for the preparation, approval and testing purposes. The preparation and approval datasets have been utilized in the preparation period of medication arrangement, and the testing dataset is utilized for the confirmation of our proposed framework and the proposed EFPN. The preparation and approval datasets remember more than 3800 pictures for every classification, with one single medication contained in each picture. Then again, the testing dataset covers 1193 pictures with 8490 explanations, where each picture incorporates 5-8 distinct medications arbitrarily chose from the 612 classifications. Tests of the DPID are displayed in Figs. 5a-c.



Fig. 5 Examples of DPID

(a) Samples with rotation factor on training and validation datasets, (b) Example with regard to changes in lighting and shot angles, (c) Samples of multiple drugs on testing dataset

## B. Implementation details

The medication pills recognition framework comprises of two phases: the localisation stage and the grouping stage. In the preparation period of the localisation stage, the names are either object (having a place with the forefront) or not (part of the foundation). For an obscure medication picture, the medication not entirely set in stone by the Origin ResNet v2 in the arrangement stage. The characterization stage centers around the order task with an enormous number of medication classes. The proposed framework incorporates two CNN models, and the boundaries of the models are inconsequential. For the situation where both the shapes and shades of the medications are shrouded in the medication localisation model, the medication localisation model needn't bothe with to be retrained when the quantity of medication classes for drug arrangement is expanded.

The designs are all start to finish prepared by two NVIDIA 1080Ti GPU, and the Adam optimiser is utilized for the inclination plunge. Four pictures have been utilized in the scaled down clump with an underlying learning pace of 10<sup>-4</sup>. The learning rate can be changed by a worth of 10<sup>-1</sup> when the exactness rates have not superior in excess of five ages during the approval stage. The proposed framework works under Ubuntu 16.04 LST with Python adaptation 3.6.4. Tensorflow-gpu 1.5.1 and Keras 2.1.6 have been utilized for the preparation and testing stages.

## C. Evaluation and discussion

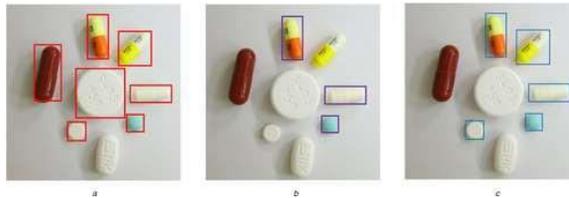
We have utilized the preparation, approval and testing datasets for the assessment of the proposed framework. The preparation and approval datasets have been utilized for the assessment of the classifier models, which are utilized during the preparation and approval stages. The testing dataset has been utilized for the medications localisation and the total framework assessment.

In the medications localisation stage, the plan of the EFPN depends on the FPN combined with the GCN, and ResNets-50 has been picked as the spine. To additionally lessen the computational intricacy, the P2 layer has been taken out. The three layers, P3, P4, and P5 of the FPN have been held and two layers, P6, and P7 have been created by the GCN to further develop the items localisation capacity. The examinations between the EFPN and the FPN, as well as the correlations between the spine ResNet-50 and the ResNet-101, are displayed in Table 2. The FPN-ResNet-50 and the FPN-ResNet-101 have been executed for drugs localisation with the right paces of 48 and 76%, separately. Figs. 6a-c show that the bigger articles are hard to situate by both the FPN-ResNet-50 and the FPN-ResNet-101. Notwithstanding, the GCN can really give the open field to the medications localisation task. The identified aftereffects of the proposed EFPN-ResNet-50 and the EFPN-ResNet-101 are 96.3 and 96.5%, separately. The exhibitions of the EFPN-ResNet-101 and the EFPN-ResNet-50 are comparative, yet the computational intricacy of the ResNet-101 is higher than the ResNet-50. The outcomes show that the EFPN has accomplished great execution on the medications localisation task. In

this manner, the ResNet-50 has been picked as the EFPN spine.

**Table 2.** Comparison of the EFPN and FPN on drugs localisation task.

| Backbone              | Pyramid layers              | Annotation (detected/totals) | Correct rate, % |
|-----------------------|-----------------------------|------------------------------|-----------------|
| FPN-ResNet-50         | {P2, P3, P4, P5}            | 4093/8490                    | 48.2            |
| FPN-ResNet-101        | {P2, P3, P4, P5}            | 6403/8490                    | 75.5            |
| <b>EFPN-ResNet-50</b> | <b>{P3, P4, P5, P6, P7}</b> | <b>8164/8490</b>             | <b>96.3</b>     |
| EFPN-ResNet-101       | {P3, P4, P5, P6, P7}        | 8180/8490                    | 96.5            |



**Fig. 6** Comparison of drugs localisation with difference detection models

**(a)** EFPN-ResNet-50, **(b)** FPN-ResNet-50, **(c)** FPN-ResNet-101

Since the ResNet is the foundation of the FPN, we have additionally assessed the ResNet-50 and the ResNet-101 classifiers in drug characterization. The blunder paces of the ResNet-50 and the ResNet-101 out of 200 ages are displayed in Table 3. To give the mistake rate bend in various ages, the graphically assessed aftereffects of the ResNet-50 and the ResNet-101 are given in Figs. 7a and b. Both of the diagrams have a similar way of behaving with the error rates meeting after 200 ages. In any case, the exhibition of the ResNet series is shaky. The total framework with the ResNet has been assessed on the approval dataset, and the blunder paces of Top-1, Top-3, and Top-5 are shown in Table 4. The Main 1 mistake rates of the ResNet-50 and the ResNet-101 are 43.5 and 39.7%, individually. In spite of the fact that the EFPN-ResNet-50 models perform well in drugs localisation, the similar appearances of the medications in different classes and the revolution factors still lead to the unfortunate acknowledgment rate in drug arrangement. To manage this problem, the proposed framework has two phases, the localisation stage, and the classification stage. The main stage centers around the areas of the drugs within the first picture. The subsequent stage attempts to perceive the medication category from the first image

**Table 3.** Error evaluation of CNN classification models with 200 epochs on single drug classification

| Network models/epoch | 20    | 40    | 60    | 80    | 100   | 200   |
|----------------------|-------|-------|-------|-------|-------|-------|
| VGG-16 [30]          | 77.0% | 29.7% | 18.3% | 11.3% | 9.3%  | 4.4%  |
| VGG-19 [30]          | 85.1% | 57.6% | 27.7% | 16.6% | 13.8  | 5.1%  |
| ResNet-50 [31]       | 69.1% | 52.6% | 27.4% | 21.4% | 44.9% | 13.4% |
| ResNet-101 [31]      | 79.0% | 48.9% | 33.7% | 22.0% | 23.8% | 16.3% |
| Inception V3 [33]    | 48.2% | 35.6% | 26.5% | 16.4% | 11.9% | 2.7%  |
| Inception V4 [35]    | 22.0% | 9.9%  | 6.6%  | 3.8%  | 3.4%  | 2.0%  |

The proposed framework has been developed utilizing the EFPN and the CNN models. The decision of classifiers is a significant issue in the proposed framework. A few CNN model have been assessed with 300 ages in Fig. 7c and d including the VGG-16/19[[33]], the Origin V3 [[37]], the Commencement V4 [[38]], the Xception [[12]], and the Initiation ResNet v1/v2 [[38]]. Table 3 records the error rates of the CNN models for the approval of single medication characterization. The results gave in Fig. 7 and Table 3 show that the VGG and the ResNet both reach combination after 200 ages, yet the bends are still in the non-steady state. Conversely, the Xception, the Origin V4, and the Beginning ResNet v2 have quicker assembly in 40-60 ages. As per the different convolution kernels used to build the portion width of classifier units, the Inception series accomplishes better exhibitions in the element extraction of the drug content.

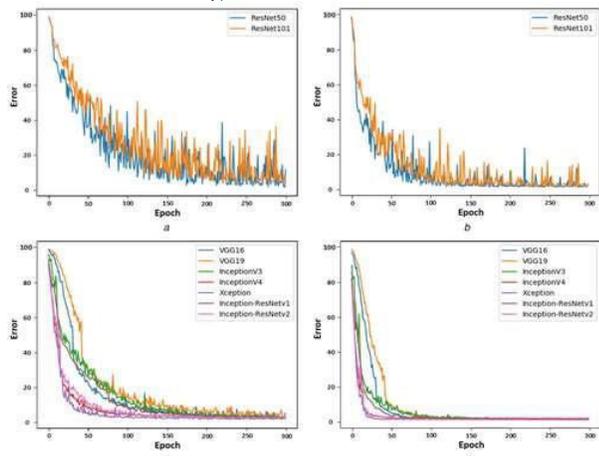


Fig. 7 [Open in figure viewer](#) [PowerPoint](#)

Error rate curve of signal drug classification with 300 epochs

(a), (b) Top-1 and Top-5 error rate curves of ResNet-50 and ResNet-101, (c), (d) Top-1 and Top-5 error rate curves of CNN-based classifiers

Table 4. Comparison with different classifiers of a complete system with top-1, top-3, and top-5 error evaluations

| Network models           | Top-1 error | Top-3 error | Top-5 error |
|--------------------------|-------------|-------------|-------------|
| EFPN-VGG 16              | 59.5%       | 54.9%       | 48.2%       |
| EFPN-VGG 19              | 55.7%       | 49.7%       | 43.5%       |
| EFPN-ResNet-50           | 43.5%       | 34.7%       | 28.2%       |
| EFPN-ResNet-101          | 39.7%       | 31.0%       | 26.5%       |
| EFPN-Inception v3        | 23.4%       | 13.1%       | 9.3%        |
| EFPN-Inception v4        | 20.9%       | 11.4%       | 9.4%        |
| EFPN-Xception            | 18.7%       | 9.5%        | 6.2%        |
| EFPN-Inception-ResNet v1 | 19.1%       | 11.8%       | 8.9%        |
| EFPN-Inception-ResNet v2 | 17.9%       | 7.6%        | 5.3%        |

#### D. Time complexity analysis

The time-cost estimation has been utilized in the time intricacy examination. Two CNN models, the ResNet and the Beginning ResNet v2, have been utilized for the medication localisation task and the medication characterization task, individually. Each picture contains 0-8 medications for the time-cost estimation. The time-cost estimations of the medication localisation task and the medication grouping task are examined

independently. In the medication localisation task, the quantity of anchors to be determined is similar in each picture and the midpoints finding opportunity is 65 ms. The medication pictures have been edited for drug order through anticipated facilitates from the medication localisation task. In the medication grouping task, the typical time for drug arrangement is 26.7 ms. The time-cost of the grouping task increments relatively as the quantity of medications in each time step. The time intricacy examination is displayed in Table 5.

Table 5. Time complexity analysis of the proposed system with images of 0–8 drugs

|                | 0    | 1    | 2     | 3     | 4     | 5     | 6     | 7     | 8     | Avg.  |
|----------------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| proposal       | 65.1 | 93.6 | 117.5 | 148.6 | 174.2 | 197.3 | 224.2 | 251.3 | 278.3 | 172.2 |
| localisation   | 65.1 | 65.7 | 64.9  | 64.9  | 64.8  | 64.9  | 65.0  | 64.9  | 65.1  | 65.0  |
| classification | 0.0  | 27.8 | 52.5  | 83.7  | 109.4 | 132.1 | 159.1 | 186.4 | 213.2 | 26.7  |

a The average time is the classification time of each drug. Unit: millisecond (ms).

Albeit the proposed framework performs well in the approval step of the preparation stage, the outcomes got from the testing stage aren't exactly great as the preparation stage. The incredibly high comparability or the comparative appearance of the medications is as yet the justification for mis-distinguishing proof. Instances of wrong acknowledgments are displayed in Fig. 9. In Fig. 9a, the pills have been delivered by a similar drug industrial facility with a similar producer's blemish on one side and a comparative engraving code on the opposite side. In Fig. 9b, the pills have comparative tones with next to no engraving code on one or the other side. In Fig. 9c, the pills have a line engraved in the center and an engraving code on the two sides of the line engrave. In Fig. 9d, the pills have an engraving code on one side with comparative tones and shapes.

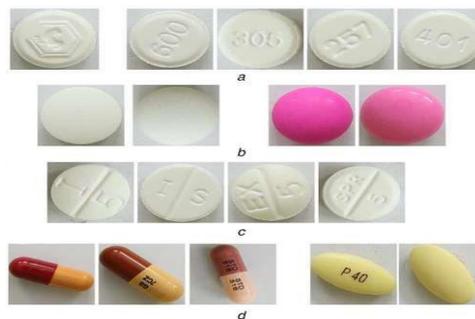


Fig. 9 Examples of erroneous recognitions

(a) Different pills have same manufacturer's mark and similar imprint code, (b) Different pills have similar colours and without any imprint code, (c) Similar colours and imprints on different pills, (d) Similar colours and shapes on different pills

## RESULT

In this paper, a programmed drug pills discovery framework has been proposed for different medication location. The proposed framework tackles two issues in the medication pills location for pragmatic applications, specifically the various pills identification and the arbitrarily positioned pills recognition. The proposed EFPN is utilized in various medication localisation, and the CNN classifiers, Beginning ResNet v2, and a medication picture data set are utilized to recognize haphazardly positioned drug pills.

Two phases have been executed in the proposed framework and each stage contains a freely prepared CNN model. The proposed EFPN is first used to find the medication areas no matter what the shapes and shades of the medications; the localisation model spotlights on finding the medication's area no matter what the grouping task. Accordingly, the localisation model needn't bother with to be retrained by and large. The Origin ResNet v2, which focuses on multi-classification arrangement errands, has been decided for drug grouping. Also, we have developed the DPID, which incorporates 612 classifications of medications for profound learning research. As per the trial results, the EFPN can accurately distinguish 96.3% of the medication areas for the approval dataset in the DPID. The exactness paces of the proposed framework, the EFPN-Beginning ResNet v2, are 82.1, 92.4, and 94.7% for rankings of Top-1, Top-3, and Top-5, individually. There are three fundamental objectives for our future work: the proceeded with progress of the medication identification precision, the rearrangements of the two-CNN model design, and the extension of the DPID to cover a greater amount of the professionally prescribed drugs.

## ACKNOWLEDGEMENT

The satisfaction and euphoria accompanying the successful completion of any task are inseparable from acknowledging the invaluable contributions of those who made it possible. Success, epitomizing hard work and perseverance, finds its steadfast support in encouraging guidance. With heartfelt gratitude, we extend our sincerest appreciation to our project guide, Mr. Mahesh Kini M, Senior Associate Professor in the Department of Computer Science & Engineering, whose unwavering inspiration and valuable guidance illuminated our path. His enthusiastic mentorship and critical reviews have been instrumental in our journey. Moreover, the selection and timely completion of this project owe much to the interest and persuasion of our project coordinator, Mrs. Vidya, Senior Assistant Professor in the Department of Computer Science & Engineering, whose contribution we will forever cherish. We express our profound gratitude to Dr. Manjunath Kotari, Professor and Head of the Department of Computer Science & Engineering, whose relentless support served as the driving force behind the project's completion. Additionally, we extend our thanks to our beloved Principal, Dr. Peter Fernandes, for his unwavering assistance and support throughout. Our deep appreciation also goes to the Management of Alva's Institute of Engineering and Technology, Mijar, Moodbidri, for providing an enabling environment conducive to the project's success. Lastly, we acknowledge the invaluable

assistance rendered by all the teaching and non-teaching staff of the Department of Computer Science & Engineering.

## REFERENCES

- [1] Ushizima, D., Carneiro, A., Souza, M., *et al*: 'Investigating pill recognition methods for a new national library of medicine image dataset'. *Int. Symp. on Visual Computing*, Las Vegas, NV, USA, December 2015, pp. 410–419
- [2] Bekker, C., Gardarsdottir, H., Egberts, A., *et al*: 'Unused medicines returned to community pharmacy: an analysis of medication waste and possibilities for redispensing', *Int. J. Clin. Pharm.*, 2017, **39**, (1), pp. 240–240
- [3] Lenzer, J.: 'US could recycle 10 million unused prescription drugs a year, report says', *Br. Med. J.*, 2014, **349**, pp. 1–1
- [4] National Health Service, U.K.: 'Pharmaceutical waste reduction in the NHS' (NHS England, England, 2015), pp. 1–24
- [5] Neto, M.A.V., Souza, J.A.W.M.D., Filho, P.P.R.C., *et al*: 'Cofordes: an invariant feature extractor for the drug pill identification'. *IEEE 31st Int. Symp. on Computer-Based Medical Systems*, Karlstad, Sweden, June 2018, pp. 30–35
- [6] Zeng, X., Cao, K., Zhang, M.: 'Mobiledeppill: a small-footprint mobile deep learning system for recognizing unconstrained pill images'. *Proc. of the 15th Annual Int. Conf. on Mobile Systems, Applications, and Services*, Niagara Falls, NY, USA, June 2018, pp. 56–67
- [7] Wang, Y., Ribera, J., Liu, C., *et al*: 'Pill recognition using minimal labeled data'. *IEEE Third Int. Conf. on Multimedia Big Data*, Laguna Hills, CA, USA, April 2017, pp. 346–353
- [8] Yu, J., Chen, Z., Kamata, S.-I., *et al*: 'Accurate system for automatic pill recognition using imprint information', *IET Image Process.*, 2015, **9**, (12), pp. 1039–1047
- [9] Chen, R.-C., Chan, Y.-K., Chen, Y.-H., *et al*: 'An automatic drug image identification system based on multiple image features and dynamic weights', *Int. J. Innov. Comput., Inf. Control*, 2012, **8**, (5), pp. 2995–3013
- [10] Ou, Y.-Y., Tsai, A.-C., Wang, J.-F., *et al*: 'Automatic drug pills detection based on convolution neural network'. *Int. Conf. on Orange Technologies*, Bali, Indonesia, October 2018, pp. 1–4

- [11] Lin, T.-Y., Dollár, P., Girshick, R., *et al.*: ‘Feature pyramid networks for object detection’. *IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017, pp. 936–944
- [12] Chollet, F.: ‘Xception: deep learning with depthwise separable convolutions’. *IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017, pp. 1800–1807
- [13] Peng, C., Zhang, X., Yu, G., *et al.*: ‘Large kernel matters – improve semantic segmentation by global convolutional network’. *IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, June 2017, pp. 1743–1751
- [14] Hartl, A.: ‘Computer-vision based pharmaceutical pill recognition on mobile phones’. *14th Central European Seminar on Computer Graphics*, Budmerice, Slovakia, May 2010, pp. 51–58
- [15] Lee, Y.-B., Park, U., Jain, A.K.: ‘Pill-ID: matching and retrieval of drug pill imprint images’. *20th Int. Conf. on Pattern Recognition*, Istanbul, Turkey, August 2010, pp. 2632–2635
- [16] Caban, J.J., Rosebrock, A., Yoo, T.S.: ‘Automatic identification of prescription drugs using shape distribution models’. *19th IEEE Int. Conf. on Image Processing*, Orlando, FL, USA, October 2012, pp. 1005–1008
- [17] Chen, Z., Kamata, S.-I.: ‘A new accurate pill recognition system using imprint information’. *Sixth Int. Conf. on Machine Vision*, London, UK, December 2013, pp. 1–8
- [18] Yu, J., Chen, Z., Kamata, S.-I.: ‘Pill recognition using imprint information by two-step sampling distance sets’. *22nd Int. Conf. on Pattern Recognition*, Stockholm, Sweden, August 2014, pp. 3156–3161
- [19] Suntronsuk, S., Ratanotayanon, S.: ‘Automatic text imprint analysis from pill images’. *9th Int. Conf. on Knowledge and Smart Technology*, Chonburi, Thailand, February 2017, pp. 288–293
- [20] Suntronsuk, S., Ratanotayanon, S.: ‘Pill image binarization for detecting text imprints’. *13th Int. Joint Conf. on Computer Science and Software Engineering*, Khon Kaen, Thailand, July 2016, pp. 1–6
- [21] Yaniv, Z., Faruque, J., Howe, S., *et al.*: ‘The national library of medicine pill image recognition challenge: an initial report’. *IEEE Applied Imagery Pattern Recognition Workshop*, Washington, DC, USA, October 2016, pp. 1–9
- [22] Girshick, R., Donahue, J., Darrell, T., *et al.*: ‘Rich feature hierarchies for accurate object detection and semantic segmentation’. *IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014, pp. 580–587
- [23] He, K., Zhang, X., Ren, S., *et al.*: ‘Spatial pyramid pooling in deep convolutional networks for visual recognition’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, **37**, (9), pp. 1904–1916
- [24] Girshick, R.: ‘Fast R-CNN’. *IEEE Int. Conf. on Computer Vision*, Washington, DC, USA, December 2015, pp. 1440–1448
- [25] Ren, S., He, K., Girshick, R., *et al.*: ‘Faster r-cnn: towards real-time object detection with region proposal networks’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, **39**, (6), pp. 1137–1149
- [26] He, K., Gkioxari, G., Dollár, P., *et al.*: ‘Mask r-cnn’. *IEEE Int. Conf. on Computer Vision*, Venice, Italy, October 2017, pp. 2980–2988
- [27] Zhang, X., Yuan, Y., Wang, Q.: ‘ROI-wise reverse reweighting network for road marking detection’. *29th British Machine Vision Conf.*, Newcastle, UK, 3–6 September 2018, pp. 1–12
- [28] Li, K., Cheng, G., Bu, S., *et al.*: ‘Rotation-insensitive and context-augmented object detection in remote sensing images’, *IEEE Trans. Geosci. Remote Sens.*, 2018, **56**, (4), pp. 2337–2348
- [29] Zhang, S., Wen, L., Bian, X., *et al.*: ‘Single-shot refinement neural network for object detection’. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018, pp. 4203–4212
- [30] Yuan, Y., Xiong, Z., Wang, Q.: ‘VSSA-NET: vertical spatial sequence attention network for traffic sign detection’. *IEEE Transactions on Image Processing*, 2019, pp. 3423–3434
- [31] Cheng, G., Zhou, P., Han, J.: ‘Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images’, *IEEE Trans. Geosci. Remote Sens.*, 2016, **54**, (12), pp. 7405–7415
- [32] Cheng, G., Han, J., Zhou, P., *et al.*: ‘Learning rotation-invariant and fisher discriminative convolutional neural networks for object.
- [33] Simonyan, K., Zisserman, A.: ‘Very deep convolutional networks for large-scale image recognition’. *Int. Conf. on Learning Representations*, San Diego, CA, USA, May 2015, pp. 1–14

[34] He, K., Zhang, X., Ren, S., *et al*: ‘ Deep residual learning for image recognition’. *IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016, pp. 770–778

[35] Neubeck, A., Gool, L.V.: ‘ Efficient non-maximum suppression’. *18th Int. Conf. on Pattern Recognition*, Hong Kong, China, August 2006, pp. 850–855

[36] Lin, T.-Y., Goyal, P., Girshick, R., *et al*: ‘ Focal loss for dense object detection’. *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, 2017

[37] Szegedy, C., Vanhoucke, V., Ioffe, S., *et al*: ‘ Rethinking the inception architecture for computer vision’. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, July 2016, pp. 2818–2826

[38] Szegedy, C., Ioffe, S., Vanhoucke, V., *et al*: ‘ Inception-v4, inception-resnet and the impact of residual connections on learning’. *Proc. of the Thirty-First AAAI Conf. on Artificial Intelligence*, San Francisco, CA, USA, February 2017, pp. 4278–4284