Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777103

AN EPISODIC APPROACH BASED ON TEXT DETECTION IN SPORTS IMAGES

Maruthupandi J. [1]*, **Riddhica S.** [2], **Haridha K** [3], **Kavi Priya J.** [4]

Affiliation:1,2,3,4- Department of Information Technology, Mepco Schlenk Engineering College, Tamil Nadu, India

*Corresponding author- Department of Information Technology
Mepco Schlenk Engineering College, Tamil Nadu, India
Email: maruthupandi@mepcoeng.ac.in

Abstract- The proliferation of sports-related multimedia content on the internet has presented interesting research challenges for effective visual search and retrieval. The challenges include poor image quality, a wide range of possible camera angles, pose variations of athletes, text deformations on clothing in motion, and occlusions caused by other objects. To overcome these challenges, a new method for identifying text present on the human body in sports-related images. is proposed in this paper. This method for detecting text in images differs from existing approaches that rely on identifying a player's torso, face, and skin. Instead, it uses an integarted episodic learning approach with inductive learning criteria to identify clothing regions within the image. The process involves combining a Residual Network with a Pyramidal Pooling Module to produce a spatial attention map, and the Progressive Scalable Expansion Algorithm is reformed for text detection from these regions. The experimental results on several benchmarks indicate that the suggested method surpasses existing methods in F1-score and precision. Outcomes on sports images from detection of text nature scene datasets such as CTW1500 and RBNR also show that across various inputs, the method that has been suggested is both dependable and efficient.

Index Terms— Clothing detection, residual network, region proposal network, text detection.

INTRODUCTION

The utilization of multimedia content, particularly the sharing of sports information at anytime and anywhere over a range of devices,

exponentially rises when communication and pooling, and progressive scaling expansion networks into a single architecture.This is, to the best of our knowledge, the first study using episodic learning for this application. We use deformable clothing areas to increase text detection accuracy, as opposed to existing methods that use faces, torsos, and skin. When compared to current approaches, the suggested method is new since it integrates spatial information provided by ResNet, regions of interest derived by PPM, and character shapes provided by PSENet[4]. Finally, we created a brand-new dataset that is excellent for this use.

RELATED WORK

As text identification in sports photographs is related to text detection in natural scenes, we compare the approaches for nature scenes, marathon images, and sports images. Long et al.[3] suggested a versatile architecture for detecting text in natural scene photos. The analysis regards text instances as sequences that are ordered and identifies symmetric axes by taking into account radius and orientation data.The approach may function well when the text is brief and just comprises a few characters. In the case of sporting photos, the existence of a single character or number, as well as brief names, is widespread. For text detection in natural scene images, Wang et al.[4] suggested a Progressive Scale Expansion Network (PSENet).

This method employs segmentation-based detectors to anticipate numerous text occurrences. However, the strategy may not work effectively for short messages with few characters. For natural scene images, Ma et al.[10] suggested a rotation region suggestion network. It considers region of interest rotation as pooling layers for the categorizer. For text with multiple characters, angular or directional information is useful otherwise, direction may produce inaccurate results. Feng et al.[11] suggested a method for identifying text in natural scene images that may be orientated freely.The method use the RoI-Slide operator to link a succession of quadrangular sentences. The approach is based on the concept of Long et al’s
The technique prioritizes on improving text detection performance by fixing tight boundary boxes for any type of text. When points are extracted from irregularly shaped characters in sports images due to non-rigid clothing, the method's performance may degrade. Text identification from multi-view natural scene images was proposed by Wang et al.[22] It seeks correspondence between several viewpoints in order to get solutions. The approach finds connection by estimating similarity and dissimilarity between text components in many perspectives. The technique necessitates many perspectives.

Most approaches for text detection in natural scene images discussed above leverage character information's direction and aspect ratio for developing deep learning networks to meet the challenge of arbitrary orientation. It was discovered that none of the techniques included garment information in sports photographs for text recognition. Words of short length relative to text in natural scene photographs, single digit numerals, partial occlusion owing to garment folding, distortion, and body motions can be expected in the case of clothing-contained text. As a result, character’s true structures may be lost. As a result, natural scene text identification approaches may be inadequate for dealing with the issues of sports photographs. Some strategies are offered to leverage multimodal concepts like face, skin, torso, and human body parts information for attaining better outcomes in sports photographs to lessen the complexity of the difficulties in sports images.

Ami et al.[23] suggested utilising facial information to recognise language in marathon photos. The technique identifies the face first, followed by the body, which carries bib numbers for detection. As long as the face is discernible in the pictures, the method operates effectively.It does not otherwise. Shivakumara et al.[5] offered torso segmentation without facial information for bib number identification in marathon images to solve this constraint. However, the technologies described above only the identify text in thoracic areas and not in other portions of the human body. To enhance the identification of bib numbers and text in sports internet technologies see rapid expansion. Users will rather get pertinent information in this situation than bulk data from a vast volume of data. For instance, scoring fours and sixes in cricket, or reaching a specific goal in soccer, etc[1]. In order to study their errors and identify weaknesses in their opponents, cricket bowlers and batsmen must also have access to specialized films. Data must be appropriately annotated at the semantic level in order to retrieve pertinent information based on user interests [2]. Text on clothes plays an important part in recovering needed information because it gives information that is close to the content of video images when identifying players or tracing marathon runners in film for the purpose of indexing and retrieval.

Furthermore, multi-modal approaches that use torso, skin, and face information reduce task complexity by accurately detecting face, skin, and torso. There is a substantial risk of losing and missing face, body, or skin information due to large fluctuations in camera perspectives, position variations, and occlusions. This observation encourages us to propose cloth detection as a type of context evidence for text detection in this study, as opposed to torso, skin, and face detection. This is due to the fact that garment detection is resistant to the aforementioned difficulties [3] and is unaffected by them. Furthermore, it is important to observe that text is commonly present on the uniform or jersey of each player in sports photos, serving as a unique identifier and providing contextual information.

A novel integrated episodic learning approach for identifying words in sports photographs has been introduced as the key contribution of this study, which incorporates area proposal networks and residual, pyramidal

method for retrieving quadrangle instances. It may not function well for short texts, which are typical in spot pictures, because it is dependent on text directions. Baek et al.[12] suggested a text identification approach based on character awareness in natural scene photos. It uses character relationships to recognize text in photos. It is unclear, that the process works for single characters. RaghuNandan et al.[13] suggested a text detection approach for natural scene, video, and born digital photos. For finding text candidates, the approach employs bit plane and convex deficit ideas. The approach does not work effectively if a character skips a few pixels owing to a cluttered background. Xu et al.[14] suggested a technique for detecting irregular scene text in photos of natural scenes. To establish the bounding box of any orientation, it ascertains the correlation among the text in question and the adjacent border. However, the approach has not been tested on sports graphics with single letters or digits. For text identification in natural scene images, Cai et al.[15] suggested an Inside-to-Outside Supervision network. It creates a hierarchical supervision module to capture texts with diverse aspect ratios, and with multiple scale supervision.

As it is widely prevalent for enhancing object segmentation by predicting correct masks, the Mask-R-CNN[16],[17] has been researched for text identification in natural scene photos. In their study, Lyu et al. [18] proposed a neural network that is capable of integrated training for detecting text in photographs of nature scenes.This method creates object form masks and recognizes text by segmenting instance areas. This approach might be unsuitable for brief text string observations .For text detection in nature scene images, Wang et al.[20] presented a quadrilateral scene text detector. To get better outcomes, the technique is divided into two steps. However, the quadrilateral suggestion network may be ineffective for random oriented and irregular text. Based on an adaptable bezier curve network, Liu et al.[21] intended a method for text extraction in natural

scene images.

photographs. Rather of relining just on torso areas, Nag et al.[24] devised a technique for detecting human body components. However, the method's efficacy is dependent on the identification of human body components. Kamlesh et al. [6] exploited text information in marathon photographs to re-identify people. For human re-identification, the technique identifies text in the race photographs. The approach works effectively for high-quality photos but not for low-quality photographs, according to testing data the utilization of multi-modal approaches has tackled certain text detection difficulties in sports images, with the effectiveness of these methods relying on pre-processing procedures like identifying faces, torsos, and other human body parts. Furthermore, lettering is frequently visible on the uniform or jersey, but not on all portions of the human body in sports images.

For recognising text in sports images, the approaches neglect clothing information. This discovery encourages us to employ garment information as context in this study to recognise text in sports photographs. This analysis reveals that none of these systems have investigated episodic learning for detecting text in sports images. Most text detection algorithms incorporate torso, face, and skin information. In the event of deformations, these approaches may not be as robust for text detection. It is only logical to attempt to identify the apparel in such photographs since the fascinating text in sports images frequently appears on clothing.

The major benefit of episodic learning is that the suggested approach may be taught with dataset samples different than the testing dataset to get the desired results. In this investigation, we use examples from different datasets focused on clothing and text detection to perform clothing and text detection, but not the datasets used for assessment. As a result, the suggested solution is capable of handling the issues on deformable clothing area.

To extract the aforementioned observations, we propose integrating Residual Network (ResNet), Pyramid Pooling Module (PPM), and Progressive Scalable Expansion Network (PSENet) for text identification on human body in sports photographs regardless of harmful impacts deformable areas. The ResNet is used to extract spatial relationships between clothing pixels in order to identify relevant information in photographs. ResNet alone will not collect deformable garment information because to the cluttered background and insufficient quality. In comparison to state-of-the-art approaches, in Fig.1 and Fig.2 the ResNet, PPM and PSENet is depicted for text detection in sports images.

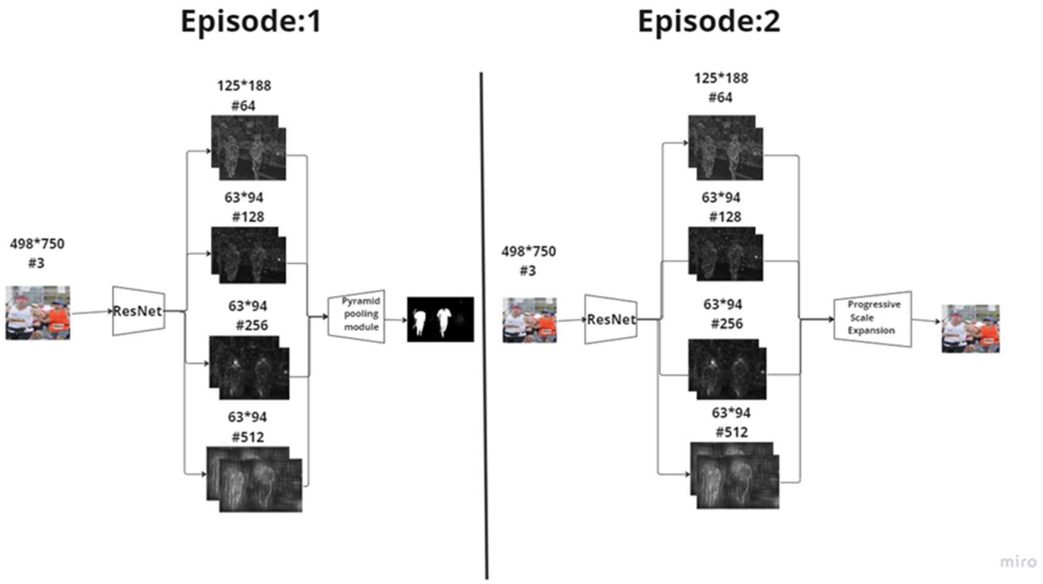


Fig.1 Network architecture as training episodes

III . METHODOLOGY

As previously stated, the attire worn by individuals in sports images, such as uniforms or jerseys, is a crucial element that includes distinct identifiers, names, and advertising materials for each input image.If we can detect clothing regions, we can detect image regions of interest as well .Therefore, we regard the clothing region as a contextual characteristic for detecting text in sports images. Inspired by deep's huge success we use Residual Network (ResNet) learning models for because it can generate strong semantic charactersistics by defining the spatial relationship of clothing pixels in place of traditional networks such as VGG-16. It is important to acknowledge that detecting text on
sports player's clothing poses challenges due to variations in poses, deformable attire, different of channels, {H , h1, h2, h3} represents height, while{W ,w1, w2, w3} represents width of output feature maps. The spatial attention network, denoted as, F_atten produces a heatmap where every element of the attention map corresponds to the clothing region and has values closer to 1, while the other elements of the attention map that are not related to the clothing region have values close to 0. From the method[26],where pyramid pooling module has been utilized for addressing the issue of varying font size texts. In this method feature map φ from previous layers are fed to 4-level parallel pooling layers which further generates the feature maps of dimensions {RCx1x1, RCx2x2, RCx3x3, RCx6x6}. In this study, we favor average pooling over max pooling since it yields better results.Our proposed method performs 1x1 convolution layer to reduce the dimension to 512 output channels that consists of ReLU activation ,batch normalization layer, and a bilinear interpolation layer for each output of adaptive pooling layer. Finally they are concatenated to a single future map. It follows the same as above to aquire the attention map π. The loss function to train F_CNN also with the F_atten is defined as :

$$L_{atten}= L_{NLLLoss}(\pi,\hat{\pi}) + \lambda_{\,dsup}\,(\pi^{*},\hat{\pi}) \qquad (1)$$

Where π ̂denotes the ground truth of regions containing cloth, L_NLLLoss is the negative log likelihood loss function and 〖λ" " 〗_dsup is a hyper parameter whose value can be fixed.

B. Text Detection From Clothes Area Proposed Network

The system described in the prior segment's network's features, 1, 2, 3 are sent to the(RPN) Region Proposal Network F_RPN for text detection. Inspired by the method in [4], where a sectioning based approach was employed to enhance text identification performance, we propose using the similar approach for isolating text instances by employing a kernel-based architecture termed Progressive Scale Expansion Network (PSENet). The loss function is constructed for PSENet training according to Equation (2). camera
perspectives, and the existence of singular digits or characters.

ResNet is insufficient to meet the aforementioned challenges. As a result, motivated by the special property of PPM that extracts regions of interest, we investigate PPM for strengthening the ResNet features which are utilized to achieve precise clothing detection, and the Progressive Scale Expansion Network (PSENet) is utilized to analyze the contours of objects and characters, regardless of their colour, size, or typeface. This attribute prompted us to investigate PSENet for text identification from garment areas identified by the combination of ResNet and PPM. The suggested approach employs an episodic training process that combines ResNet, PPM, and PSENet features for word identification in sports photos.

A.Attention ResNet for Human Cloth Detetion:

It is a CNN Backbone to extract a 3DConvolutional feature map as φ ϵR^(c,h,w), φ1 ϵR^(c1,h1,w1),φ2 ϵR^(c2,h2,w2),φ3 ϵR^(c3,h3,w3).
Here φ is the output generated by the final residual block of CNN backbone, , φ1 is the output of the second last block of CNN backbone, φ2 is the third last, and φ3 is the fourth last. {C, c1, c2, c3} represents numbers

$$F_{dice}(S_i, G_{i,}) = \frac{2\Sigma_{x,y}(S_{i,x,y}, X\, G_{i,x,y})}{\Sigma_{x,y}S^2_{i,x,y} + \Sigma_{x,y}G^2_{i,x,y}} \quad (2)$$

where Lc stands for the loss of a complete text instance and Ls for the loss of a shrunk text instance, The dice coefficient is represented by F_dice. The value of pixel (x, y) in the n segmentation result is Si of PSENet, Gi,x,y denotes pixel values of (x,y) in ground truth.

C. Training and evaluating Spatial Attention and Region Proposal Networks using episodic methods.

To train clothing detection features and text detection features from clothing regions, we suggest Episodic training [27], a text detection network that is integrated for sports images.

The suggested approach is comprised of two well structured training episodes. { F_CNN , F_atten } using Latten and {F_CNN , F_RPN } using L_RPN that exposes F_CNN to a different information (i.e., either the information of F_atten or F_RPN) during each training episodes. While F_atten to obtain the desired results of capturing the clothing region of a human, F_RPN is employed while is used to estimate the text regions in the input image. As F_CNN performs feature extraction for both 〖{F〗_CNN , F_atten }the output feature maps {ψ, ψ1, ψ2, ψ3} gain robustness and possess the ability to capture information related to both clothing and text. We employ DeepFashion2 Dataset [28], which offers a instance mask for several types of garment regions, to train {F_CNN , F_atten } utilising Latten. The suggested approach groups all the various clothing types into a single class.

In L_atten , a binary mask is utilised as the ground truth to determine whether or not a zone is a garment region. The ICDAR 2015 [29] dataset is used to train the proposed technique. since it offers bounding boxes for texts, which serve as the basis for LRPN. Algorithm-1 presents the Episodic Training algorithmic steps. Table I contains a description of the variables used in the algorithm-1.

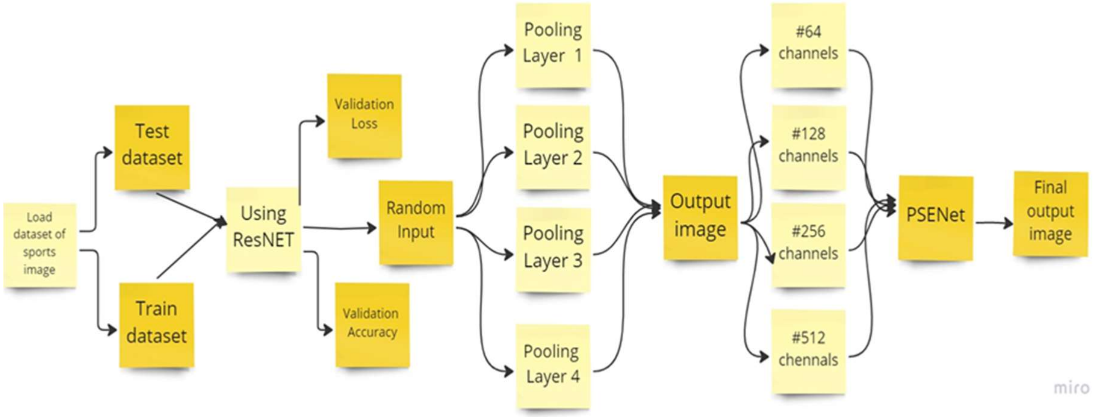


Fig 2. Network Architecture

The suggested method employs algorithm-1 to train the {F_CNN , F_atten , F_RPN } and then, without further training, merges the three parts into a novel mechanism. As seen in Figure 1, the output from the F_CNN is sent to 〖F〗_atten , which applies a deformation to a spatial attention map to create a two-dimensional matrix.

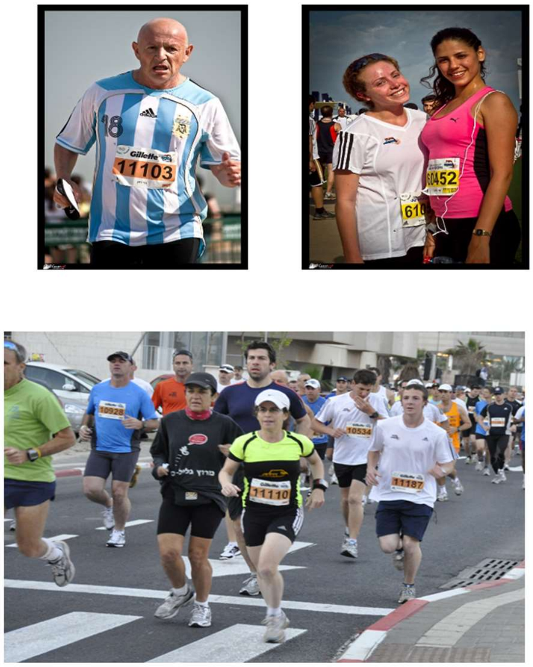


Fig.3 Sample images of Datasets

We resize π to have shape of the output feature map from F_CNN using bilinear interpolation.Then we perform Hadamard product to combine all the channels of {Ψ, Ψ1, Ψ2, Ψ3,} and to infer a new feature map {Ψ*, Ψ*1, Ψ*2,} Ψ*3, like defined in Equation (3). Our method feeds the new {Ψ*, Ψ*1, Ψ*2,Ψ*3,} to F_RPN for spotting text occurences.

$$\begin{aligned} \Psi^* &= \pi \otimes \Psi \\ \Psi^*1 &= \pi \otimes \Psi 1 \qquad (3) \\ \Psi^*2 &= \pi \otimes \Psi 2 \end{aligned}$$

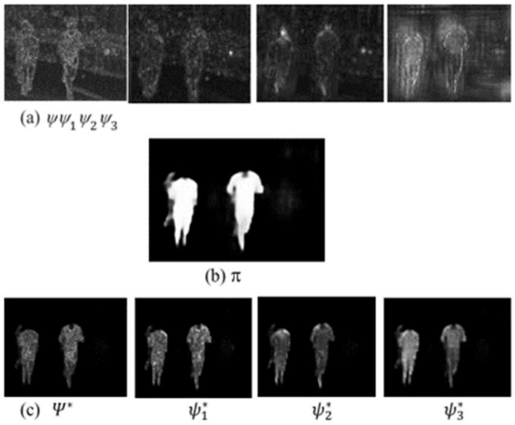
$$\Psi*3 = \pi\otimes \Psi3$$

Where $\otimes$ denotes pointwise multiplication within each channel."

The suggested technique trains on samples from the DeepFashion2 dataset [28] for apparel detection and the ICDAR 2015 dataset [29] for text detection.To clarify, for training and testing in our study, we adhere to a technique for across dataset validation. Consequently, we refrain from utilizing any training dataset samples are utilized during the testing phase of the method.

That may be shown from the results in Fig. 3, where one can see how the flexible garment region affects the text. These instances show how the approach functions.

The principal goal is to abstain traditional multistep procedures for text detection by omitting extraneous supporting details. This causes an RPN network to be very dependent on a previous module. The proposed approach, on the other hand, gets around this restriction by using an attention mechanism. In other words, we can infer that Fatten serves as a guiding signal for F_RPN , causing it to concentrate on garment regions as opposed to the backdrop, as seen in Fig. 4, which presents the outcomes of the backbone's intermediate phases (F_CNN). As a result, the pixels are sharpened where there is clothing by suppressing other pixels, as demonstrated in Figure 4(a). Figure 4(b) illustrates the impact of the attention network (ResNet-16 + PPM), where these regions are clearly discernible. This is demonstrated in Figure 4(c) of the corresponding intermediate results (a).


(a) $\psi\psi_1\psi_2\psi_3$

(b) π

(c) Ψ^* ψ_1^* ψ_2^* ψ_3^*

Fig: 4 spatial attention network

TABLE I

THE LIST OF DISTRIBUTION ELEMENTS UTILIZED IN THE ALGORITHM-1

Variable	Definition
Ψ	result of the final residual block of CNN backbone
$\psi1$	result from the second final residual block of CNN backbone
$\psi2$	result from the third final residual block of CNN backbone
$\psi3$	result from the fourth final residual block of CNN backbone

C,c_1, c_2, c_3	Number of Channels
H', h_1, h_2, h_3	Represents Height
W', w_1, w_2, w_3	Represents Width of feature maps
π	Attention map
L_c	Loss for full text instance
L_s	Loss for reduced ones
β	Learning rate
$\hat{\pi}$	Ground truth of containing cloth region
$L_{NLLLoss}$	Negative Log Likelihood to detect ground truth clothing region
λ_{dsup}	Hyper parameters

IV. EXPERIMENTAL RESULTS

In sports photographs, we have chosen to test our methods using sports images from the benchmark nature environment datasets of CTW1500 and RBNR [23].

A. Dataset Preparation and Analysis

We build a dataset by compiling photographs from many sources, including the internet, YouTube, soccer, tennis, cricket, marathons, and our own collections. The photos in this dataset include shots with altered poses, camera angles, and target-to-camera distances. Text on non-rigid materials like uniforms or jerseys can cause perspective distortion in photos because of the aforementioned issues. Sports photographs also include bib numbers for the athletes as well as scene texts. These variances make the dataset more complex.

Similar to this, we use two common datasets, RBNR [23],and CTW1500 [19], which provide marathon photos with prominently displayed bib numbers, to exhibit the strength of the suggested approach. Similarly, in order to showcase the effectiveness of our approach, we acquire images from benchmark datasets of natural scenes that showcase both human bodies and text information namely, CTW1500 [19] and RBNR Text [23].

TABLE II
INFORMATION ON VARIOUS DATASETS CONSIDERED FOR EXPERIMENTATION

Datasets	Number of Images	
	Train	Test
RBNR Data-Marathon	150	67
CTW-1500	350	151
Our Dataset	500	300

The sports images picked from these two datasets are difficult to achieve high results compared to Kamlesh's method [6] utilizes text detection and recognition to identify the person in the natural scene images but not specifically for sports images. Table II provides more information about

images. We also use the recently established approach [24] for jersey number recognition in sports images. It performs a preprocessing phase called human body part is detection to provide results. We use the most recent deep learning based methods for comparative studies, namely The method proposed by Long et al [3], TextSnake, which offers a flexible representation for detecting text in natural scene images, and the method of Wang et al,[4], which proposes PESNet to address issues with comparing pixel-wise segmentation methods for text detection in sports images with those designed for natural scene images might not be effective for sports images. In the same way, we contrast our method with the Lyu et al. [18] method, which was created for natural scene photos, to demonstrate that the approaches that use Mask-R-CNN for text detection are ineffective in the situation of deformed garment regions. These three approaches should be taken into consideration for comparative studies since they deal with issues including arbitrary orientation, complicated backgrounds, low contrast, and low resolution that are typical for sports photos. In this study, we use labelled samples to empirically calculate the following values for episodic training. The evaluation experiments in this work are conducted using a consistent setup and values. The learning rate of 0.02 is set using a stochastic gradient descent optimizer for all experiments in this work over a total of 31122 iterations. Following a second session of training to improve FCNN, FRPN for Ntext = 72 iterations, FCNN, Fatten is trained for Ncloth = 73 iterations. The learning rate's decay is set to 0.0003. The value of λpse is set to 0.5, and λdsup is give a value of 0.4. The current algorithms were trained using samples from the same dataset that was utilised for testing in experiments on all datasets, including our own.

B. Ablation Research:

In our method, as discussed in the proposed methodology section, spatial attention on clothing is generated using ResNet-18 in this work and

combinations, indicates that ResNet-18 is capable of balancing the need to detect text occurrences[31] with the need to generate fewer false positives. Overall, we can confirm that PPM, RestNet and PSPNet, are capable of handling the difficulties associated with text identification[32] in sports images. As the fundamental stage of the proposed model is the uniform dress code or jersey detection, sample qualitative findings of the proposed model on clothing detection for pictures of various datasets. The proposed method show that the suggested combination of PPM+ResNet-18+ PSE performs well for photos of various types of deformable clothing.

C. DarkNet For Accurate Text Detection:

In our paper we have detected the text in sports images and bounded by a boundary using PSENet Text Detector. It has a disadvantage that it does not detect the text properly and display the contents.So we have used DarkNet method and YOLO for text detetction in sports images. Darknet is an open-source neural network framework that can be used for object detection, including text detection. Some advantages of Darknet over PSENet for text detection are faster processing time, accurate detection and easy to use. Along with DarkNet framework YOLO algorithm is used for accurate text detection. YOLO (You Only Look Once) is a popular object detection algorithm that can be used in a variety of applications. Some common usages of YOLO are object detection in images. YOLO can detect objects in images with high accuracy and speed, making it useful for applications such as automated surveillance, self-driving cars, and image search engines. Object detection in videos. YOLO can detect objects in videos in real-time, making it useful for applications such as video surveillance, traffic monitoring, and sports analysis.

YOLO can be trained to detect faces in images and videos, making it useful for applications such as face recognition and authentication. Overall, YOLO is a versatile algorithm that can be applied in many different fields and industries. In the context of text detection for sports images, YOLO

is an object detection algorithm that can be used

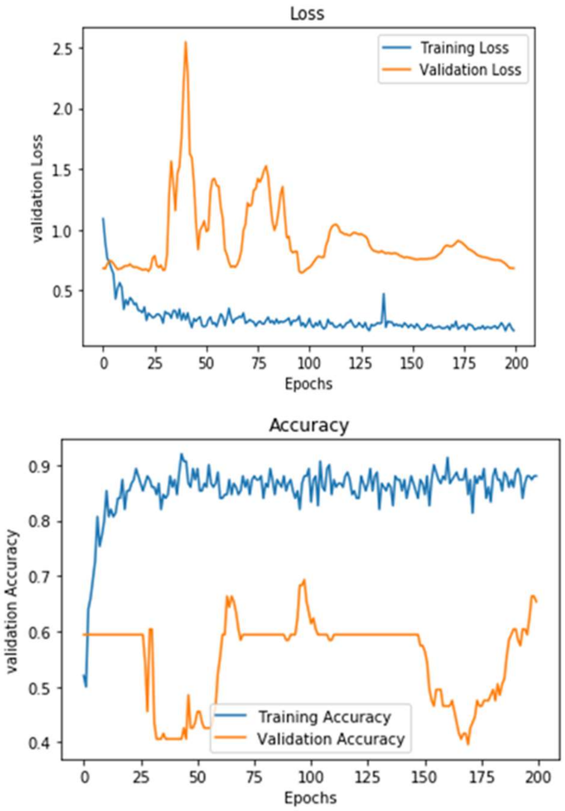


Fig.6 Training Accuracy and Validation Accuracy

V. CONCLUSION AND FUTURE WORK

We suggest a new approach to detecting text in sports images using episodic learning. Our proposed architecture combines residual networks and region proposal networks into a single framework that can detect clothing with a uniform dress code and jerseys. It then applies PSENet to detect text. In contrast to other methods that utilize features such as facial expressions, skin colour, torso shape, and other body parts of humans, to minimize the complexity of background in sports images, our method uses deformable clothing regions to detect text. The reason for using deformable clothing regions for one challenge in detecting text in sports images is that occlusions can cause body parts such as the face, skin, and torso to vanish., while clothing is less likely to disappear entirely.

Our proposed model uses Episodic training to combine features extracted from utilizing ResNet for the segmentation of clothing areas and PSENet of RPN for text detection. Our experiments on various datasets, including marathon runners and natural scene datasets, demonstrate that our technique exhibits superior performance to existing techniques in precision and F-measure metrics. No previous research has employed an athlete's apparel for the purpose of text detection prior to this study. However, our method is not entirely reliable in challenging cases, which we aim to investigate in future research.

REFERENCES

[1] C. Xu, Y.-F. Zhang, G. Zhu, Y. Rui, H. Lu, and Q. Huang, “Using webcast text for semantic event detection in broadcast sports video,” IEEE Trans. Multimedia, vol. 10, no. 7, pp. 1342–1355, Nov. 2008.

[2] S. Zhang, J. Huang, H. Li, and D. N. Metaxas, “Automatic image annotation and retrieval using group sparsity,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 42, no. 3, pp. 838–849, Jun. 2012.

[3] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, “Textsnake: A flexible representation for detecting text of arbitrary shapes,” in Proc. ECCV, 2018, pp. 19–35.

[4] W. Wang et al., “Shape robust text detection with progressive scale expansion network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 9328–9337.

[5] P. Shivakumara, R. Raghavendra, L. Qin, K. B. Raja, T. Lu, and U. Pal, “A new multi-modal approach to bib number/text detection and recognition in Marathon images,” Pattern Recognit., vol. 61, pp. 479–491, Jan. 2017.

[6] P. Xu, Y. Yang, and Y. Xu, “Person re-identification with end-to-end scene text recognition,” in Proc. CCCV, 2017, pp. 363–374.

[7] B. Bataineh, S. N. H. S. Abdullah, and K. Omar, “A novel statistical feature extraction method for textual images: Optical font recognition,” Expert Syst. Appl., vol. 39, no. 5, pp. 5470–5477, Apr. 2012.

[8] S. M. S. Ismail, S. N. H. S. Abdullah, and F. Fauzi, “Detecting and recognition via adaptive binarization and fuzzy clustering,” J. Sci. Technol., vol. 27, no. 4, pp. 1759–1781, 2019.
network for spotting text with arbitrary shapes,” in Proc. ECCV, 2018, pp. 71–78.

[9] D. Li, Z. Zhang, X. Chen, and K. Huang, “A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios,” IEEE Trans. Image Process., vol. 28, no. 4, pp. 1575–1590, Apr. 2019.

[10] J. Ma et al., “Arbitrary-oriented scene text detection via rotation proposals,” IEEE Trans. Multimedia, vol. 20, no. 11, pp. 3111–3122, Nov. 2018.

[11] W. Feng, W. He, F. Yin, X. Y. Zhang, and C. L. Liu, “TextDragon: An end-to-end framework for arbitrary shaped text spotting,” in Proc. ICCV, 2019, pp. 9076–9084.

[12] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character region awareness for text detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 9365–9374.

[13] K. S. Raghunandan, P. Shivakumara, S. Roy, G. H. Kumar, U. Pal, and T. Lu, “Multi-Script-oriented text detection and recognition in video/scene/born digital images,” IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 4, pp. 1145–1162, Apr. 2019.

[14] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang, and X. Bai, “TextField: Learning a deep direction field for irregular scene text detection,” IEEE Trans. Image Process., vol. 28, no. 11, pp. 5566–5579, Nov. 2019.

[15] Y. Cai, W. Wang, Y. Chen, and Q. Ye, “IOS-net: An inside-to-outside supervision network for scale robust text detection in the wild,” Pattern Recognit., vol. 103, Jul. 2020, Art. no. 107304.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proc. ICCV, 2017, pp. 2980–2988.

[17] Z. Huang, Z. Zhong, L. Sun, and Q. Huo, “Mask R-CNN with pyramid attention network for scene text detection,” in Proc. ICCV, 2019, pp. 764–772.

[18] P. Lyu, M. Liao, C. Yao, W. Wu, and X. Bai, “Mask textspotter: An end-to-end trainable neural
Song, and T. M. Hospedales, “Episodic training for domain generalization,” in Proc. ICCV, 2019, pp. 1446–1455.

[19] S. Roy, P. Shivakumara, U. Pal, T. Lu, and G. H. Kumar, “Delaunay triangulation based text detection from multi-view images of natural scene,” Pattern Recognit. Lett., vol. 129, pp. 92–100, Jan. 2020.

[20] S. Wang, Y. Liu, Z. He, Y. Wang, and Z. Tang, “A quadrilateral scene text detector with two-stage network architecture,” Pattern Recognit., vol. 102, Jun. 2020, Art. no. 107230.

[21] Y. Liu, H. Chen, C. Shen, T. He, L. Jin, and L. Wang, “ABCNet: Realtime scene text spotting with adaptive bezier-curve networ,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 9809–9818.

[22] C. Wang, H. Fu, L. Yang, and X. Cao, “Text co-detection in multi-view scene,” IEEE Trans. Image Process., vol. 29, pp. 4627–4642, 2020.

[23] I. B. Ami, T. Basha, and S. Avidan, “Racing bib number recognition,” in Proc. BMCV, 2012, pp. 1–12.

[24] S.Nag, R. Rmachandra, P. Shivakumara,

U. Pal, T. Lu, and M. Kankanhalli, “CRNN based jersey number/text recognition in sports and Marathon images,” in Proc. ICDAR, 2019, pp. 1149–1156.

[25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. CVPR, 2016, pp. 770–778.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[27] D. Li, J. Zhang, Y. Yang, C. Liu, Y. Z.

[28] Y. Ge, R. Zhang, L. Wu, X. Wang, X. Tang, and P. Luo, “A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing ing images,” in Proc. CVPR, 2019, pp. 5337–5345.

[29] D. Karatzas et al., “ICDAR 2015 competition on robust reading,” in Proc. ICDAR, 2015, pp. 1156–1160.

[30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. CVPR, 2017, pp. 6230–6239.

[31] Maruthupandi.J, Vimala Devi. K “Efficient Feature Extraction for Text Mining” January 2016. Advances in Natural and Applied Sciences 10 (4)pp. 64-73.

[32] Shalini Vincent M, Maruthupandi. J, Vimala Devi. K, (2015), “A Novel Fuzzy Based Clustering For Multi-Label Text Categorization”,
International Journal of Applied Engineering Research, ISSN 0973-4562, Vol. 10 No.2 (2015)pp. 1642-1647.

.