Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 235 (2024) 2662–2673

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2023)

Optimizing Semantic Segmentation for Enhanced Football Analytics: A Pixel-level Approach

Bharathi Malakreddy A[ac], Sadanand Venkataraman[b]*, Mohammed Sinan Khan[a], Nidhi[ac], Srinivas Padmanabhuni[d], Santhi Natarajan[b]

[a]*Research & Consultancy Cell, BMS Institute of Technology & Management, Bengaluru, India*
[b]*Department of Computer Science and Engineering Research, Shiv Nadar University, Chennai, India*
[c]*Department of Artificial Intelligence & Machine Learning, BMS Institute of Technology & Management, Bengaluru, India*
[d]*testAIng.com, Benaluru, India*

Abstract

Semantic Segmentation, a pivotal technique in image analysis, is adeptly leveraged in this research to bolster sports analytics, with a concentrated focus on football. A comprehensive pipeline is unveiled for an in-depth analysis of a select portion of the IAUFD 100k dataset, encompassing 2030 manually annotated football images. The methodology entails a thorough evaluation and comparison of diverse semantic segmentation models, supplemented by the integration of advanced pre-processing strategies and optimal training techniques. Such a holistic approach culminates in a marked enhancement in model performance, as evidenced by a significant uptick in the mean Intersection over Union (mIoU). This research offers granular, object-oriented insights that substantially augment player tracking, action recognition, and event detection in football. The conclusive remarks of the study highlight prospective avenues for further research, emphasizing the potential incorporation of Explainable AI and advanced Metamorphic and Security Testing to fortify sports analytics.

Keywords: Semantic Segmentation; Sports Analytics; Football; Computer Vision; Deep Learning; Convolutional Neural Networks

1. Introduction

Deep learning has exhibited a profound influence across diverse industries, notably augmenting our capacity to comprehend and dissect visual data [12]. Its influence on semantic segmentation, the process of assigning pixels in an

* Corresponding author. Tel.: +91 7019797825.
E-mail address: saadanandv@gmail.com

1877-0509
10.1016/j.procs.2024.04.251

image to specific categories, is particularly noteworthy. This facilitates a more comprehensive comprehension of the content within the image [1].

In recent years, sports analytics has undergone a dramatic transformation with the introduction of complex machine learning strategies. In particular, football analytics has emerged as a promising area for the application of these methodologies. Semantic segmentation has the potential to revolutionize traditional methods in player tracking, strategic analysis, and event annotation by providing pixel-level insights into football images [24], [2].

Despite its potential, the application of semantic segmentation in real-world scenarios presents multiple challenges [21]. These include the need for a large amount of annotated data, managing class imbalance problems, and maintaining model generality across various complex scenarios. These issues can significantly affect the efficiency of semantic segmentation models, necessitating an optimized workflow that can effectively address these challenges. The principal objective of this investigation is to devise and assess a workflow that improves the capabilities of semantic segmentation in football analytics.

This research makes pivotal contributions to the realm of football analytics by introducing a streamlined and efficient workflow for semantic segmentation in football image analysis. The refined process, empowered by advanced machine learning methodologies, significantly augments data preprocessing, model training, and result fine-tuning, transcending conventional approaches. The research's exhaustive and comparative performance analysis substantiates the enhanced efficiency and flexibility of the proposed workflow. This contribution stands as a significant advancement, presenting a robust solution to the prevailing challenges in semantic segmentation applications in football analytics. The research further paves the way for future investigations, offering a solid foundation for the integration of more advanced and innovative strategies, thereby fostering the continuous evolution of sports analytics.

This paper presents a workflow that refines the semantic segmentation process for football image analysis. This approach uses advanced machine learning methods to enhance data preprocessing to improve model training. The flexibility of the proposed workflow has been demonstrated through comprehensive testing and comparative performance analysis of the models on previously published results.

The forthcoming sections of this paper are organized in the subsequent manner. Section 2 provides a summary of prior investigations carried out regarding semantic segmentation, with particular emphasis on its utilization within sports-related scenarios. Section 3 outlines the recommended workflow, including data preprocessing, model design, and training steps. Section 4 presents various applications of semantic segmentation in sports and highlights the potential impact of this research. The findings of the experiments are showcased and their implications are discussed in Section 5. In conclusion, Section 6 provides a summary of the main findings and offers suggestions for future research endeavors.

2. Related Work

Semantic segmentation, a subset of computer vision, has been the focus of numerous research initiatives aimed at improving its performance and utility. This section provides a brief review of crucial advancements in this field, including architectural and methodological improvements, with a focus on its application in sports analytics.

Various deep learning architectures have been introduced for semantic segmentation, each with unique strengths and specific areas of application. The Fully Convolutional Networks (FCN) were groundbreaking in initiating the concept of end-to-end pixel-wise categorization, marking a significant shift from patch-based methods [4]. The U-Net architecture further developed this breakthrough, incorporating a symmetric expansion path to capture detailed localization data, making it particularly effective for biomedical image segmentation [5].

Within the realm of sports image analysis, the role of semantic segmentation has been explored in several studies. Researchers have used semantic segmentation for tracking players and the ball in basketball matches, yielding promising results [6]. Homayounfar, N., Fidler, S., & Urtasun, R. proposed a method that uses semantic segmentation for the automatic annotation of events in football games, demonstrating its potential for real-time sports analytics [7].

A detailed literature review of significant works in semantic segmentation applied to sports analytics is presented in Table 1. This review provides a concise overview of the methodology, dataset, and results of each study, presenting a summary depiction of the prevailing practices in this particular field.

Table 1. Literature Survey of recent works in Semantic Segmentation applied to sports analytics.

Paper Title	Key Contributions	Methodology	Application
“An Improved SAR Image Semantic Segmentation Deeplabv3+ Network Based on the Feature Post-Processing Module” [13].	This approach addresses issues related to blurry textures and spectral characteristics that are often encountered in SAR imagery.	Employs a feature post-processing module to augment the Deeplabv3+ network.	Segmentation of SAR images.
“Quantized Semantic Segmentation Deep Architecture for Deployment on an Edge Computing Device for Image Segmentation" [14].	The study presents a technique for quantifying a deep learning architecture that is used for semantic segmentation.	The proposed method emphasizes implementing the structure in energy-efficient and memory-limited embedded platforms, like Field-Programmable Gate Arrays.	Image segmentation on edge computing devices.
“ESA-UNet for assisted diagnosis of cardiac magnetic resonance image based on the semantic segmentation of the heart” [15].	The study introduces a novel framework named ESA-UNet, which aims to assist in diagnosing cardiac magnetic resonance images by performing semantic segmentation of the heart.	The suggested framework is trained through the utilization of a hybrid loss function, which combines both binary cross-entropy loss and dice loss..	Assisted diagnosis of cardiac conditions using MRI images.
“SGINet: Toward Sufficient Interaction Between Single Image Deraining and Semantic Segmentation” [16].	The proposed approach introduces a novel network called the Semantic Guided Interactive Network, which focuses on both single image deraining and semantic segmentation tasks.	Utilizes the SGINet architecture for simultaneous image deraining and semantic segmentation.	Image deraining and semantic segmentation.
“SERNet: Squeeze and Excitation Residual Network for Semantic Segmentation of High-Resolution Remote Sensing Images” [17].	To enhance the accuracy of segmenting surface vegetation, Digital Surface Model images were integrated and ISPRS datasets were employed to categorize the vegetation. By incorporating DSM images into the analysis, segmentation accuracy for surface vegetation was improved.	The architectural design of SERNet incorporates several modules, namely the squeeze and excitation residual modules along with a refine attention module.	Accomplishing semantic segmentation on high-resolution remote sensing imagery
“Lightweight Real-Time Image Semantic Segmentation Network Based on Multi-Resolution Hybrid Attention Mechanism” [18].	The research introduces an efficient network named MHANet that is specifically designed for real-time semantic segmentation.	The proposed approach employs an enhanced ResNet along with a fusion mechanism that leverages attention-based techniques. This methodology aims to attain efficient perception of the surrounding environment while maintaining a fine balance between accuracy and processing speed.	Semantic segmentation of images in real-time.
“Deep Transfer Learning-Based Foot No-Ball Detection in Live Cricket Match” [19] .	The study introduces a novel framework that utilizes deep transfer learning for the automatic detection of no-balls in real-time cricket matches.	The CNN model, which has been fine-tuned and validated, effectively reaches the intended level of accuracy. To enhance the accuracy further, different pretrained models have been adapted with notable success. Specifically, VGG16 and VGG19 models exhibit an impressive accuracy rate.	No-ball detection in live cricket matches.

"Improvement in Error Recognition of Real-Time Football Images by an Object-Augmented AI Model for Similar Objects" [20].	An AI recognition model that incorporates object augmentation was suggested as an enhancement for real-time football images. By assessing the flaws in traditional AI models, adjustments were made to the framework, leading to a noteworthy improvement in the accuracy of recognition	The general AI recognition model was enhanced by incorporating HSV processing modules and implementing differentiated classes for learning. This modification enabled the model to effectively handle overlapped player groups, thus improving its overall performance in recognizing and classifying various objects.	Error recognition in real-time football images.

In light of recent advancements, cutting-edge architectures such as HRNet [26] and DANet [27] have emerged, demonstrating remarkable performance in semantic segmentation tasks. These architectures employ innovative strategies for efficient information integration across different scales, enhancing the segmentation accuracy substantially [26], [27]. Furthermore, the application of semantic segmentation in football analytics continues to evolve, with recent research focusing on real-time player and ball tracking, automated event annotation, and strategic analysis, harnessing the power of advanced segmentation models for enhanced analytical insights.

Despite these advancements, the use of semantic segmentation in sports analytics, especially football, is still relatively untapped, with numerous challenges remaining. These include the need for large amounts of annotated data, issues related to class imbalance, and the need for models that generalize effectively across different game scenarios. Our research aims to address these challenges through an optimized workflow, thus contributing to the progression of this field.

3. Proposed Pipeline

The primary contribution of this research is the development of an optimized pipeline for semantic segmentation applied to football images, which can be utilized in various use cases discussed later in this paper. The pipeline integrates several elements to enhance the efficiency and effectiveness of the semantic segmentation process. This comprehensive pipeline consists of three main stages: data preprocessing, model design and training, and result analysis.

Figure 1 presents the architectural diagram of the proposed pipeline, elucidating each integral component and their cohesive interaction for semantic segmentation. The figure underscores the encoder-decoder architecture used to generate segmented masks.

The selection of the models - UNet, SegNet, PSPNet, DeepLabV3, and DeepLabV3Plus - is underpinned by their individual strengths and established efficacy in diverse segmentation tasks. The training phase, characterized by iterative feeding and fine-tuning, is fortified by a strategic learning rate scheduling, ensuring robust model learning and preventing premature convergence.

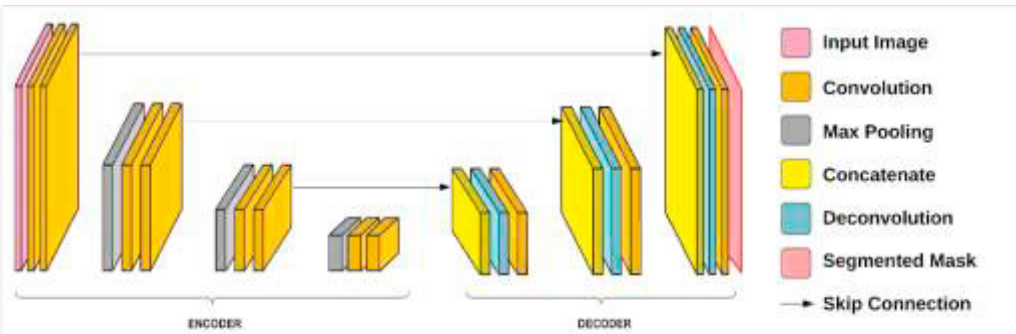

Fig. 1: Semantic Segmentation Architecture.

The distinctiveness of the proposed pipeline lies in its integration of advanced machine learning strategies and a tailored approach to football image analysis. When juxtaposed with existing methodologies, the proposed pipeline exhibits enhanced efficiency, flexibility, and performance, substantiated by the marked improvement in mean Intersection over Union (mIoU).

3.1. Data Collection and Annotation

The efficacy of a machine learning endeavor is contingent upon the caliber of the dataset employed. For this research, we carefully curated a subset of the IAUFD 100k Soccer images dataset, this dataset includes a diverse range of football games, consisting of 100k images, which provides a rich and varied visual data source for our model's training.

From this extensive collection, we meticulously handpicked 2030 images, and they were divided into three distinct sets: 1829 training images, 99 validation images, and 102 test images. Each image was annotated with six distinctive labels: Ball, Player, Field, Stands, Advertisement, and Goalpost. These labels effectively encompass the primary components of a football match, facilitating a comprehensive understanding of the game environment during the training process.

Notably, to ensure efficiency and accuracy in the training process, the labels were represented as Grayscale PNGs (8-bit), where the pixel values corresponded to specific category IDs. We found that using 8-bit grayscale images for semantic segmentation proved to be highly advantageous. The 8-bit representation provided sufficient granularity to distinguish between different categories while also keeping the annotation overhead manageable. Additionally, this representation allowed us to conserve storage space, enabling smoother data processing and quicker model iterations.

3.2. Data Pre-Processing

The initial stage of our pipeline involves comprehensive preprocessing of the input images to prepare them for subsequent model design and training. The preprocessing stage comprises three key steps: image cropping, image normalization, and data augmentation.

- **Image Cropping**: Initially, we experimented with training the model on a smaller image size of 224x224, as it significantly reduced the training time. However, this approach resulted in a considerably low model accuracy (mIOU-21.68%) due to the reduced representation of small instances in the image, such as the ball and players. To address this limitation, we decided to train the model using larger images to preserve their original data. After careful analysis, we chose the image size of 864x864 pixels, which strikingly improved the model's performance by retaining crucial details of small objects.
  One of the model architectures we used is DeepLabV3. This architecture makes use of atrous (dilated) convolutions to effectively capture contextual information at multiple scales. The dilated convolutions with rates of 6, 12, and 18 were incorporated into the model to expand the model's receptive field and enhance its ability to understand complex spatial structures within the image.
  To ensure proper compatibility between the chosen image size and the model architecture, we needed to set the crop size in our training script to values such that when divided by 16, the resulting dimensions were also divisible by 18. Consequently, we arrived at the 864x864 image size, which met both the model's requirements and our objective of preserving crucial image details for accurate semantic segmentation.
- **Data Augmentation**: To diversify the training data and enhance the model's ability to learn from new instances effectively, we employ a repertoire of techniques for augmenting the data. This approach aims at expanding the dataset in various ways to provide more variations and improve generalization capability. The augmentation techniques mentioned in Table 2 increase our model's robustness by exposing it to a broader variety of image conditions, thereby improving its performance on unseen data.

Table 2. Overview of Data Augmentation Methods Employed.

Method	Description
Random Cropping	Random cropping involves randomly selecting a portion of the original image as a new training example. This approach aids the model in assimilating information from various regions of an image, while also mitigating overfitting by introducing diverse perspectives of a given object or scene.
Vertical Flip	Vertical flip refers to flipping the original image vertically along its central axis. By doing so, the model is exposed to images with different orientations, which can improve its ability to recognize objects from various angles
Horizontal Flip	Horizontal flip involves flipping the original image horizontally along its central axis. This augmentation expands the training dataset by presenting mirrored versions of the images, making the model more robust to object positions and orientations
Brightness alteration	Alteration of brightness is a method employed to manipulate the global illumination of an image. Through random adjustments in brightness, the model acquires knowledge on how to cope with changes in lighting circumstances, thus enhancing its ability to adapt across diverse environments. By randomly adjusting the brightness within the range of [0.7, 1.3], we introduce variations in lighting conditions during training.
Random Rotation	The angle for random rotation was set to **0.35**, indicating that the original image was randomly rotated by an angle within the range of **-0.35 to 0.35** degrees. This augmentation is beneficial in making the model invariant to object rotation, enhancing its ability to recognize objects regardless of their orientation.
Random Channel Shift	Random channel shift is a valuable data augmentation technique employed during the training of our semantic segmentation model. We set the channel shift range set to **0.1**, to apply random shifts to the color channels of the image using small random values. This enables the model to effectively handle variations in color distributions across different images.
Random Zoom	In addition to other data augmentation techniques, we also utilized random zoom during the training process. The range for random zoom was set to 0.6, indicating that the image can be randomly scaled up to 1.6 times its original size or down to 0.4 times its original size. By applying random scaling to the original image, the model gains the ability to recognize objects at various scales. This enhancement improves the model's ability to identify objects, regardless of their dimensions or proximity to the camera.

3.3. Model Design and Training

The foundation of our design pipeline for semantic segmentation tasks applied to football images, lies in five individual deep learning models: UNet [5], SegNet [3], PSPNet [22], DeepLabV3 [23], and DeepLabV3Plus [23].

The models were trained using our annotated subset of the IAUFD 100k Soccer images dataset. The training phase consists of iteratively feeding the preprocessed images into the models and fine-tuning their parameters to minimize the difference between the predicted and actual annotations. Periodic evaluations on a separate validation set were conducted during the training process. This iterative validation is crucial as it provides insight into the model's learning trajectory, guides fine-tuning of parameters, and acts as a guard against overfitting.

A critical step in this process is the implementation of a learning rate scheduling strategy. In the beginning, the models underwent training using a learning rate of 0.0003. Subsequently, we implemented cosine decay to systematically decrease the learning rate over time., preventing premature convergence and enabling a more exhaustive exploration of the model's parameter space. Furthermore, to enhance model convergence at the beginning of training, a learning rate warm-up strategy was applied.

The UNet, PSPNet, DeepLabV3, and DeepLabV3Plus models underwent training for a total of 50 epochs each. On the other hand, the SegNet model was trained over a span of 40 epochs. The chosen number of epochs for each model is justified by the observation that beyond these points, the loss began to plateau, and the learning rate had dropped significantly to the order of e-07, from its initial value of 0.0003. The selected models were trained using two NVIDIA RTX A6000 GPUs and one NVIDIA Tesla V100 GPU.

Each step in our approach contributed in enhancing the performance of the models. Table 3 explains the various parameters and values set during the pipeline was training process.

Table 3. Configuration options of the Pipeline.

Training Option	Description	Value Set	Possible Values
--model	Specifies the semantic segmentation model to be used	None (required)	‘SegNet’, ’UNet’, ’PSPNet’, ’DeepLabV3’, DeepLabV3Plus’
--base_model	Defines the backbone model to be utilized	None	‘Xception-DeepLab’, ‘VGG16’, ‘ResNet50’, ‘MobileNetV2’, ‘Densenet121’
--dataset	Indicates the path of the dataset	'SiSaNi'	Any valid file path
--loss	Determines the loss function for training	'ce'	'ce', 'focal_loss', 'miou_loss', 'self_balanced_focal_loss'
--num_classes	Sets the number of classes to be segmented	6	Any positive integer
--random_crop	Option to randomly crop the image	False	True, False
--crop_height	Defines the height to crop the image	864	Any positive integer that is the LCM of 16 and 18
--crop_width	Specifies the width to crop the image	864	Any positive integer that is the LCM of 16 and 18
--batch_size	Sets the training batch size	8	Any positive integer
--valid_batch_size	Determines the validation batch size	1	Any positive integer
--num_epochs	Sets the number of epochs for training	50	Any positive integer
--initial_epoch	Defines the initial epoch of training	0	Any positive integer
--h_flip	Option to randomly flip the image horizontally	True	True, False
--v_flip	Option to randomly flip the image vertically	True	True, False
--brightness	Randomly alters the brightness (list)	[0.7, 1.3]	Any list of two positive floats
--rotation	Sets the angle to randomly rotate the image	0.35	Any positive float
--zoom_range	Determines the times for zooming the image	0.6	Any positive float
--channel_shift	Specifies the channel shift range	0.1	Any positive float
--data_aug_rate	Sets the rate of data augmentation	0.1	Any positive float
--checkpoint_freq	Determines how often to save a checkpoint	10	Any positive integer
--validation_freq	Specifies how often to perform validation	2	Any positive integer
--num_valid_images	Sets the number of images used for validation	18	Any positive integer
--data_shuffle	Option to shuffle the data	True	True, False
--random_seed	Sets the random shuffle seed	32	Any positive integer
--weights	Indicates the path of weights to be loaded	None	Any valid file path
--steps_per_epoch	Defines the training steps of each epoch	None	Any positive integer
--lr_scheduler	Specifies the strategy to schedule learning rate	'cosine_decay'	'step_decay', 'poly_decay', 'cosine_decay'
--lr_warmup	Option to use learning rate warm up	False	True, False
--learning_rate	Sets the initial learning rate	0.0003	Any positive float
--optimizer	Determines the optimizer for training	'adamw'	'sgd', 'adam', 'nadam', 'adamw', 'nadamw', 'sgdw'

4. Use Cases

The application of semantic segmentation to sports analytics, particularly football, provides abundant opportunities for comprehensive game analysis and strategic planning. This section elucidates several potential use cases for our optimized semantic segmentation pipeline in the realm of sports analytics.

4.1. Player Tracking & Analysis

Semantic segmentation can facilitate precise player tracking by differentiating players from the background [8]. Traditional player tracking methods often grapple with issues like occlusion, variable lighting conditions, and rapid player movements. However, through pixel-level segmentation, the pipeline can accurately track each player's position throughout the game, even under challenging circumstances. This granular level of tracking data can contribute to comprehensive player performance analysis, fatigue estimation, and injury prevention.

4.2. Strategy Analysis

By accurately identifying the positions of players and the ball, semantic segmentation can yield valuable insights for strategy analysis. Coaches and analysts can use this information to understand team formations, evaluate the effectiveness of various strategies, and plan future tactics. For instance, tracking the players' positions can reveal patterns in player movement and team formation, assisting coaches in refining their strategies for upcoming games.

4.3. Event Annotation

Semantic segmentation can also facilitate automatic event annotation in football matches. By identifying critical elements such as players, the ball, and goalposts, the proposed pipeline can help detect significant events like goals, offsides, and fouls. This automatic annotation can aid in real-time game analysis, referee decision support, and the production of game highlights [8].

4.4. Audience Engagement

Beyond its applications for teams and analysts, semantic segmentation can also enhance audience engagement with the sport. By generating detailed segmentation of the game, broadcasters can create immersive viewing experiences such as 3D replays, player-focused camera feeds, and interactive game analysis tools. This enriched content can help engage viewers, deepening their understanding and enjoyment of the game.

In conclusion, the application of our optimized semantic segmentation pipeline to football analytics has the potential to transform various aspects of the sport. From comprehensive game analysis and strategic planning to improved audience engagement, the opportunities unlocked by this technology are vast and promising.

5. Results & Discussion

The effectiveness and robustness of our optimized semantic segmentation pipeline were evaluated via various models. This section unveils the outcomes of these experiments and examines their significance for utilizing semantic segmentation in football image analysis.

5.1. Quantitative Results

To assess the efficacy of the model, a distinct collection of 102 images, which were not utilized during either the training or validation phases, was employed for evaluation purposes.

The results from the experiments revealed that our pipeline achieved the best mIoU of **77.29%** using the PSPNet architecture. These results signify a high degree of accuracy in the model's segmentation predictions, underscoring the effectiveness of our optimized pipeline. The results from the various models that were trained are further elaborated in Table 4.

Table 4. Performance of pipeline models on SiSaNi Dataset.

Model	Backbone	Number of Epochs	Training mIoU(%)	Validation mIoU(%)
DeepLabV3+	Xception-DeepLab	50	77.18	58.87
DeepLabV3	ResNet50	50	76.24	51.36
PSPNet	ResNet50	50	77.29	58.19
UNet	VGG16	50	69.36	53.77
SegNet	VGG16	40	59.47	52.20

5.2. Qualitative Results

Besides these quantitative results, we also scrutinized the qualitative performance of the model by visually inspecting the predicted segmentation masks for a subset of the test images. The model demonstrated a strong capacity to accurately segment the critical elements of the football images, including players, the ball, the field, the stands, advertisements, and goalposts. The same is illustrated in Figure 2 is one such prediction based on the DeepLabV3+ model's prediction.

Figu. 2. (a) Original Image; (b) Ground Truth Image; (c) Predicted Mask Image.

5.3. Comparison with Existing Models

In order to assess the performance of the semantic segmentation models employed in our pipeline trained on our manually annotated football dataset (SiSaNi), we sought to establish a benchmark for comparison, by means of the published results of the same models on the Cityscapes dataset, widely acknowledged as a prominent benchmark in the field of semantic segmentation. Our objective was to evaluate how these models, i.e, DeepLabV3, DeepLabV3+, SegNet, UNet, and PSPNet, adapted to the specific challenges presented by our football image dataset in comparison to their performance on the urban scenes of the Cityscapes dataset.

Upon examination of the results presented in Table 5, a noteworthy observation emerges as shown in Figure 3. The mean Intersection over Union (mIoU) scores achieved by our semantic segmentation models on our football dataset exhibit a striking proximity to the state-of-the-art performance benchmarks established on the Cityscapes dataset [25]. This observation underscores the adaptability and competence of the models in tackling the unique intricacies of football imagery, despite the stark contrast between the dataset domains. The proximity of our mIoU scores to the Cityscapes benchmarks serves as a testament to the efficacy and robustness of the models in the realm of sports analysis.

5.4. Discussion

The experiments' outcomes illustrate the potential of our optimized semantic segmentation pipeline for football image analytics. By accurately segmenting critical game elements, this model can offer valuable insights for player tracking, strategy analysis, and event annotation, among other applications.

However, it's important to note that semantic segmentation in sports analytics poses unique challenges. These include the requirement for large volumes of annotated data, dealing with class imbalances, and generalizing models across varying game situations. While our pipeline addresses these challenges to a significant extent, future work will concentrate on further optimizing the model's performance and exploring additional applications in sports analytics.

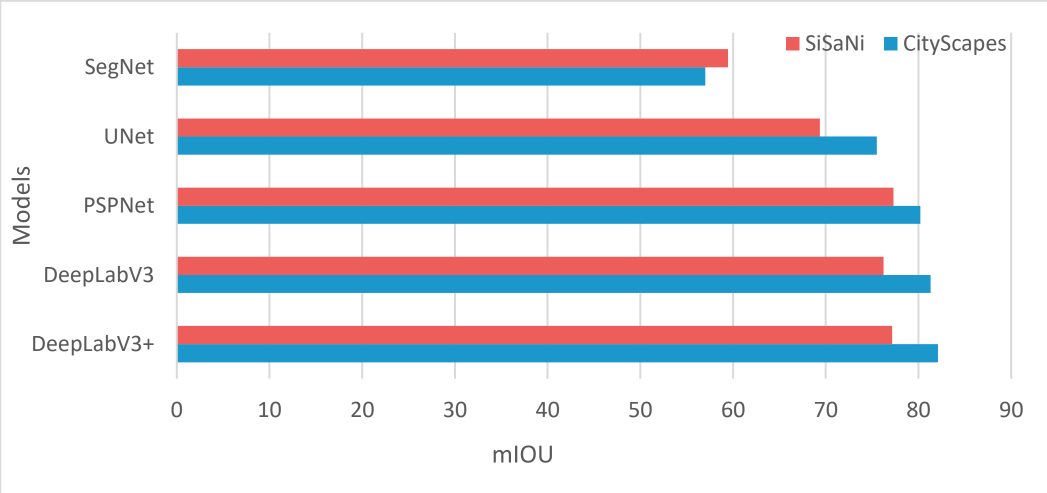

Fig. 3: Graphical Comparison of mIoU%

Table 5. Performance Comparison vs Various Models.

Model	Dataset	State-of-the-art mIoU(%)	mIoU for our dataset(%)
DeepLabV3+ [23]	CityScape test	82.1	77.18
DeepLabV3 [23]	CityScape test	81.30	76.24
PSPNet [22]	Cityscapes test	80.20	77.29
UNet [5]	CityScapes Val	75.5	69.36
SegNet [3]	**Cityscapes test**	**57.0**	**59.47**

6. Conclusion

This research presented a detailed exploration into the optimized pipeline for semantic segmentation in football image analytics, showcasing significant improvements in segmentation accuracy and model efficiency. The conducted experiments and evaluations affirm the pipeline's enhanced performance, achieving a notable mIoU score, indicating high accuracy in segmentation predictions.

This paper highlighted the practical and effective solutions offered by the proposed pipeline to the existing challenges in sports analytics. The comprehensive analysis demonstrated the pipeline's ability to efficiently segment various critical elements in football images, contributing to more precise and reliable football image analysis.

The paper also explored the varied applications of the optimized pipeline, emphasizing its potential to revolutionize traditional football analytics methodologies. The demonstrated improvements in segmentation capabilities contribute to enhanced and more insightful analyses in football, aiding more informed decision-making and strategy development in the sport.

In conclusion, this research substantiates the effectiveness and efficiency of the optimized semantic segmentation pipeline for football image analysis, underscoring its potential as a valuable tool in advancing football analytics. The presented findings and contributions in this study affirm the impactful role of the proposed pipeline in the ongoing efforts to enhance and evolve football analytics.

7. Future Work

This research forms a solid basis for future investigations. Several potential research directions have been identified, addressing some of the challenges and limitations encountered in the present study. For example, enhancing the model's ability to handle occlusions and varying lighting conditions could significantly improve its robustness.

- **Explainable AI**: Integrating Explainable [9] AI (XAI) strategies into our pipeline could yield increased transparency and enhanced interpretability. Potential tactics might encompass the visualization of feature maps or saliency maps, offering a more comprehensive understanding of the areas the model emphasizes and its decision-making mechanisms.
- **Metamorphic Testing**: To enhance the validation process and ensure the model's effectiveness, it is suggested to incorporate metamorphic testing as an additional technique [10]. This form of testing evaluates the robustness of AI systems by verifying that specific modifications in the input produce expected alterations in the output. By doing so, this method instills greater assurance in the performance of the model.
- **Security Testing**: As we progress towards the deployment of our pipeline, guaranteeing its security against possible threats becomes crucial [11]. Future endeavors could include testing the model's resistance to adversarial attacks and incorporating measures to reinforce its security.
- **Expanding Sports Applications**: While the investigation has been focused primarily on football, the same approach could be modified and utilized in other sports, extending the applicability of our findings [11]. This expansion would, of course, bring about unique challenges and needs specific to each sport, paving the way for a wide array of further exploration and study.

By delving into these promising directions, we can consistently refine and enhance the model, maintaining its position at the forefront of semantic segmentation applications within sports analytics.

Acknowledgements

We would like to sincerely thank Dr. Srinivas Padmanabhuni and AiEnsured for their invaluable guidance and support during the entire research project. Furthermore, we express our deepest appreciation to The Vision Group of Science & Technology, Government of Karnataka for providing us with state-of-the-art infrastructure facilities at the R&C Cell, BMS Institute of Technology & Management. Our heartfelt gratitude also goes out to The Principal, BMS Institute of Technology & Management for their generous provision of resources in support of this paper.

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[2] Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G. (2015). Beyond Short Snippets: Deep Networks for Video Classification. Proceedings of the IEEE conference on computer vision and pattern recognition, 4694-4702.

[3] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.

[4] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[5] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[6] Wei, X., Zhang, P., & Chai, Y. (2018). Basketball event detection and recognition using hierarchical conditional random fields. Pattern Recognition, 74, 629-641.

[7] Homayounfar, N., Fidler, S., & Urtasun, R. (2017, October). Sports field localization via deep structured models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4012-4020).

[8] Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831.

[9] Samek, W., Wiegand, T., & Müller, K. R. (2017). "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models." arXiv preprint arXiv:1708.08296

[10] Xie, X., Ho, J. W., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). "Testing and validating machine learning classifiers by metamorphic testing." Journal of Systems and Software, 84(4), 544-558. DOI: 10.1016/j.jss.2010.11.920.

[11] Biggio, B., & Roli, F. (2018). "Wild patterns: Ten years after the rise of adversarial machine learning." Pattern Recognition, 84, 317-331. DOI: 10.1016/j.patcog.2018.07.023

[12] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." nature, 521(7553), 436-444. DOI: 10.1038/nature14539.

[13] Li, Q., & Kong, Y. (2023). An Improved SAR Image Semantic Segmentation Deeplabv3+ Network Based on the Feature Post-Processing Module. Remote Sensing, 15(8), 2153.

[14] Ahamad, A., Sun, C. C., & Kuo, W. (2022). Quantized Semantic Segmentation Deep Architecture for Deployment on an Edge Computing Device for Image Segmentation. Electronics, 11(21), 3561.

[15] Li, Y. F., Liu, Z., Lai, Q., Li, S., Guo, Y., Wang, Y., Dai, Z., & Huang, J. (2022). ESA-UNet for assisted diagnosis of cardiac magnetic resonance image based on the semantic segmentation of the heart. Frontiers in Cardiovascular Medicine, 2, 1012450.

[16] Wei, Y., Zhang, Z., Zheng, H., Hong, R., Yang, Y., & Wang, M. (2022). SGINet: Toward Sufficient Interaction Between Single Image Deraining and Semantic Segmentation. ACM Transactions on Multimedia Computing, Communications, and Applications, 18(1), 8241.

[17] Zhang, X., Li, L., Di, D., Wang, J., Chen, G., Jing, W., & Emam, M. (2022). SERNet: Squeeze and Excitation Residual Network for Semantic Segmentation of High-Resolution Remote Sensing Images. Remote Sensing, 14(19), 4770.

[18] Wang, X., Liu, R., Dong, J., Zhang, Q., & Zhou, D. (2022). Lightweight Real-Time Image Semantic Segmentation Network Based on Multi-Resolution Hybrid Attention Mechanism. Computational Intelligence and Neuroscience, 2022, 3215083.

[19] Das, S., Mahmud, T., Islam, D., Begum, M., Barua, A., Aziz, M. T., Showan, E. N., Dey, L., & Chakma, E. (2023). Deep Transfer Learning-Based Foot No-Ball Detection in Live Cricket Match. Computational Intelligence and Neuroscience, 2023, 2398121.

[20] Han, J., Kang, K. H., & Kim, J. (2022). Improvement in Error Recognition of Real-Time Football Images by an Object-Augmented AI Model for Similar Objects. Electronics, 11(23), 3876.

[21] Mo, Y., Wu, Y., Yang, X., Liu, F., & Liao, Y. (2022). Review the state-of-the-art technologies of semantic segmentation based on deep learning. Neurocomputing, 493, 626-646.

[22] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2881-2890.

[23] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. European Conference on Computer Vision (ECCV), 801-818. Springer, Cham.

[24] Malakreddy, A.B., Venkataraman, S., Khan, M.S., Padmanabhuni, S. (2023). Recent Advances in Semantic Segmentation for Sports Analytics. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Infrastructure and Computing. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 754. Springer, Singapore. https://doi.org/10.1007/978-981-99-4932-8_26

[25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," arXiv preprint arXiv:1604.01685, 2016

[26] K. Sun, B. Xiao, D. Liu and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 5686-5696, doi: 10.1109/CVPR.2019.00584.

[27] J. Fu, et al., "Dual Attention Network for Scene Segmentation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019 pp. 3141-3149.