https://doi.org/10.2352/ISSN.2470-1173.2021.6.IRIACV-303

Automatic Annotation of American Football Video Footage for Game Strategy Analysis

Jacob Newman, Jian-Wei Lin, Dah-Jye Lee, and Jen-Jui Liu, Brigham Young University, Provo, Utah, USA

Abstract

Annotation and analysis of sports videos is a challenging task that, once accomplished, could provide various benefits to coaches, players, and spectators. In particular, American Football could benefit from such a system to provide assistance in statistics and game strategy analysis. Manual analysis of recorded American football game videos is a tedious and inefficient process. In this paper, as a first step to further our research for this unique application, we focus on locating and labeling individual football players from a single overhead image of a football play immediately before the play begins. A pre-trained deep learning network is used to detect and locate the players in the image. A ResNet is used to label the individual players based on their corresponding player position or formation. Our player detection and labeling algorithms obtain greater than 90% accuracy, especially for those skill positions on offense (Quarterback, Running Back, and Wide Receiver) and defense (Cornerback and Safety). Results from our preliminary studies on player detection, localization, and labeling prove the feasibility of building a complete American football strategy analysis system using artificial intelligence.

Keywords - American football, artificial intelligence, deep learning, object detection, player labeling, sports strategy analysis.

Introduction

Over the decades, the game of American football (hereafter referred to as football) has evolved to incorporate various aspects of technology with the purpose of improving both gameplay and spectating. For example, coaches in the NFL are provided with a tablet in which they can see images of player formations from earlier in the game [1]. This incorporation of technology assists the coaches in making well-informed decisions while on the field.

Computer Vision in particular has begun to play an increasingly significant role. With recent technological advancements in computer vision, tracking players in sports has become possible. For example, players in basketball [2], soccer [3], and football [4] can been tracked using a combination of computer vision and other techniques.

Motivation

One area in the field of sports that can be improved by using Computer Vision techniques is annotating sports footage. This has many similarities to tracking players while on the field. Manual labelling of individual players can be tedious and time consuming, so automating the process could prove to be beneficial.

Deep Learning provides a promising avenue of research in player tracking. Automated player detection and tracking using Computer Vision and Deep Learning could prove to be useful in helping coaches study and understand both how their team plays and how other teams play. This could assist them in game planning before the game or, if allowed, making decisions in real time, thereby improving the game.

This paper presents a method of player detection from a single image immediately before the play starts. YOLOv3, a pre-trained deep learning network, is used to detect the visible players, followed by a ResNet architecture which labels the players. This research will be a foundation for future research in tracking key players during a video. With this tracking data, automated game analysis can be accomplished.

Challenges

Various challenges exist in locating and tracking players in Football videos. One challenge arises because of the common placement of the camera. In a 45-degree camera view of the field, which is the camera view we are using, the play is seen from above and behind the line of scrimmage. This introduces the challenge of occlusion. With the camera behind the line of scrimmage, some players are easily blocked. In particular, the Quarterback often stands directly behind the Center, often prohibiting the Center from being seen. The defensive players closest to the line of scrimmage are also often occluded by the offensive line. We address this issue when labelling players with the ResNet after the visible players are detected.

Another challenge arises with the movement of the camera. The camera is not stationary through the duration of the play, which causes some players to leave the frame. This challenge will be addressed in future work. We are planning to obtain bird's eye view game videos that most teams have access for our future work. With the bird's eye view videos, these two challenges may not be so significant.

Related Work

The previous work that has been done in the field of sport player tracking has used both Computer Vision techniques and physical tracking devices, such as the RFID tracking system used by the NFL. [4] tracks the movement of football players using Computer Vision techniques to create a cross-domain transformation from a camera view to a bird's eye view. [5] predicts the trajectories of wide receivers in an football play using a Markov decision process. [6] identifies the type of formation the offensive team is lined up in by locating the line of scrimmage and using SVM classifiers.

[3] and [7] track soccer players using multiple cameras. [2] tracks basketball players using a broadcast view of the court and the play-by-play text of the game.

Using Artificial Intelligence techniques, such as Deep Learning and Machine Learning, player tracking in sports has become even more feasible. Using Machine Learning and the RFID tracking system, [8] identifies the routes of American Football plays. Also using the RFID tracking technology, [9] quantifies the Quarterback's decision making process, predicting which player will receive the pass. [10] identifies key events using Deep Learning and feature extraction. [11] uses a deep CNN to detect a soccer ball and a kalman filter to track the ball. [12] also uses Deep Learning techniques with a CNN to track soccer players.

Contributions

We present a method of locating and labelling football players using Deep Learning techniques. This eliminates the need of attaching physical hardware to the player. The sole data to be used is obtained by a single overhead camera. Applying these Deep Learning techniques presents a novel method of locating and labelling football players.

Data

The dataset used is a custom collected and labelled dataset. To assist in the collection of data, the images were obtained using the Madden NFL 20 PC game.

Data Collection

To ensure obtaining the clearest view of as many players as possible, the images collected are from the All-22 view of the field, which is above and behind the offensive line (See Fig 1). To ensure the images are usable when training, they are obtained immediately before the start of the play, when the players are on the line of scrimmage in their assigned positions. 600 images in total were collected, with 500 images used for training our model and 100 images for testing the model.

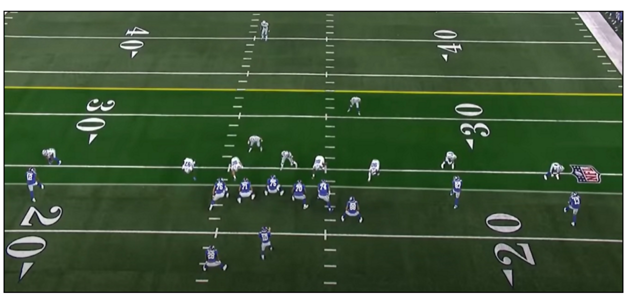


Figure 1. *Sample Image of All-22 Camera View*

Currently, because the images are obtained from a single consistent source, all images have a similar format and resolution. This currently limits the accuracy of the model when being run on outside data. However, this issue will be addressed in future work, when using data from other sources.

Data Labelling

Each of the individual players in the collected images were labelled according to their respective positions. The labelling process was done using the Microsoft VOTT software [14]. There are twelve tags in total, with seven on offense and five on defense. The offensive tags consist of the Quarterback (QB), the Running Back (RB), the Center (C), the Offensive Guard (OG), the Offensive Tackle (OT), the Tight End (TE), and the Wide Receiver (WR). The defensive tags consist of the Defensive Tackle (DT), the Defensive End (DE), the Linebacker (LB), the Cornerback (CB), and the Safety (S).

Every football play consists of twenty two players, with eleven players on offense and eleven players on defense. In order to assist in obtaining the location of all twenty two players, even partially or completely occluded ones, we labelled the occluded players with a single bounding box no larger than 5 pixels wide.

Detecting and Labelling Players

The program is split into two main sections, or modules. The purpose of the first module, which uses YOLOv3, is to identify the location of each of the visible players. The purpose of the second module, which uses a ResNet (Residual Network) based architecture, is to label each of eight groups of players. The twelve individual player positions are combined into eight groups, as follows:

- Defensive Back (DB) - Safeties and Cornerbacks
- Defensive Line (DL) - Defensive Tackles and Defensive Ends
- Linebacker (LB) - Linebackers
- Offensive Line (OL) - The Center, Offensive Guards, and Offensive Tackles
- Quarterback (QB) - The Quarterback
- Running Back (RB) - The Running Back
- Tight End (TE) - The Tight End
- Wide Receiver (WR) - The Wide Receiver

Detecting Players

The first module is based heavily on the Train Your Own YOLO Github project by AntonMu [13]. We used it as a starting point in detecting the football player's locations.

To obtain the needed data we used the Microsoft VOTT software to label the individual football players [14]. In order to obtain the highest accuracy when detecting the players, we converted each of the twelve individual player position labels into a single Player label. Doing this increased the accuracy of locating all of the players, as opposed to using all twelve tags when training.

During the Training process, we excluded all bounding boxes that are less than 5 pixels wide. This is done to avoid training the model on players that are partially or completely occluded.

After training, the results consist of the player's bounding box coordinates and confidence values. Using these results and the ground truth data of the custom dataset, we analyzed the accuracy of the YOLO network on our data.

To analyze the results for a single image, we compare the ground truth visible players to the players detected by YOLO. We compare the ground truth visible player with the closet bounding box detected by YOLO. If the center of the two bounding boxes are within 20 pixels of each other, it is considered a valid player detection. If the center of the two bounding boxes is farther than 20 pixels, it is not considered a detected player and is excluded from further processing. If, after iterating through all of the ground truth visible players, there are still YOLO detections, these are considered incorrect detections, which represent either duplicates or false positives.

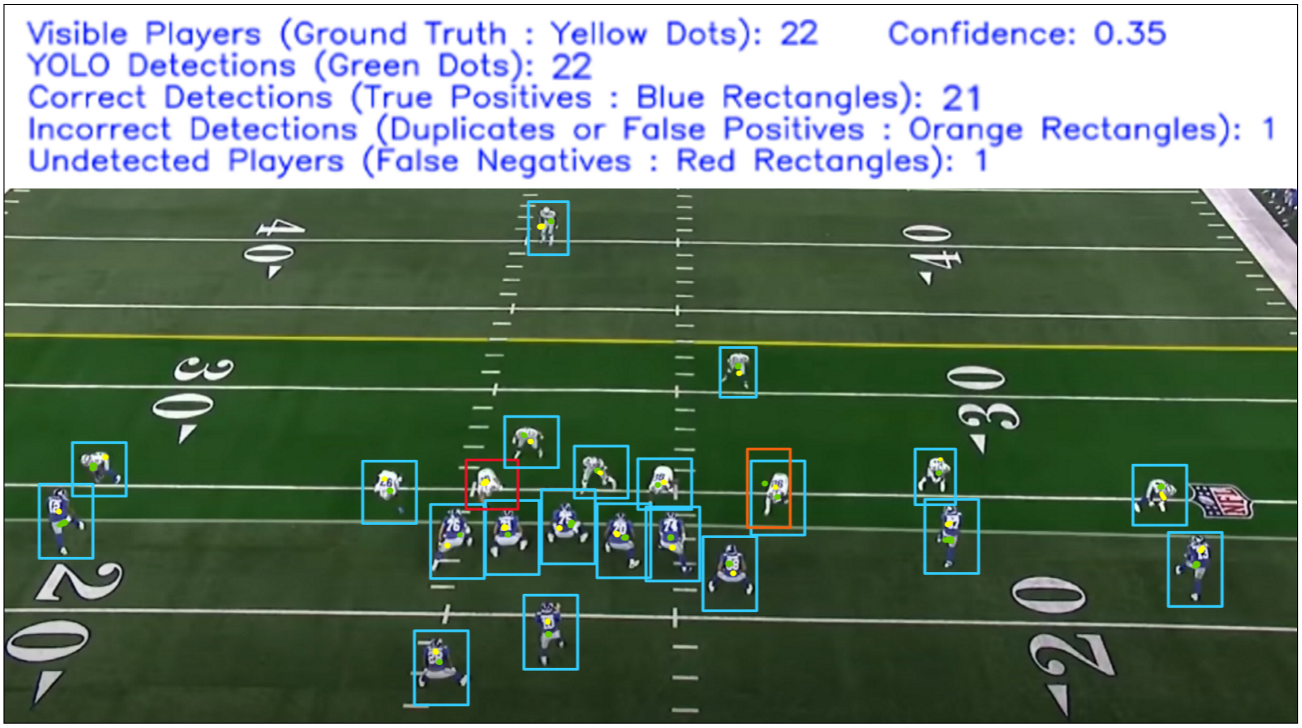


Figure 2. *Example of Image Analyzed by First Module*

Figure 2 displays an example of a graphical representation of the analysis for a single image. As seen in the analyzed image, the ground truth visible players and the players detected by YOLO are marked with a yellow and green dot, respectively. The correct detections are marked with a light blue rectangle, the incorrect detections are marked with an orange rectangle, and the players that are missed are marked with a red rectangle.

Using the results of this analysis on differing amounts of training data, we produced a PR curve. With a confidence value of 35%, our best results consisted of a Precision of 97.65% and a Recall of 91.81% using 400 training images. The results of the training on differing amounts of data is shown in Figure 3.

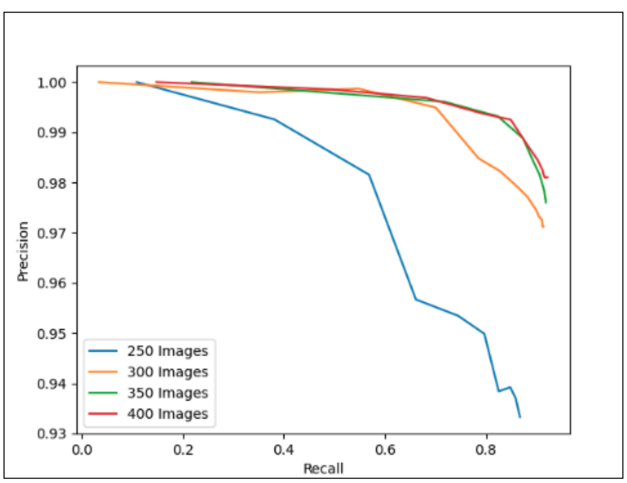


Figure 3. *Precision-Recall Results of First Module*

Labelling Player Groups

To label the player groups, a ResNet50 architecture based on work done by JK Jung is used [15]. A higher priority is placed on correctly identifying the key skill players, which consist of the Quarterback, the Running Back, and the Wide Receiver on offense, as well as the Safety and the Cornerback on defense.

ResNet and Data Processing

A graphical representation of the architecture used for the second module, which is based on a ResNet, is shown in Figures 4, 5, and 6.

To train the ResNet, we formatted the data as shown in Figure 7. Twenty two data points are created from each ground truth image collected. Cycling through each player location as a "root" location, we determine the root player (shown in yellow), the root player's team (shown in green), and the root player's opposing team (shown in light blue). This is done by creating three channels; a root channel, a team channel, and an opposing team channel. The players on offense and defense are known because of the ground truth labels. The corresponding player location(s) on the root, team, and opposing team channels are marked by a circle with a slight gradient. The gradient assists in detecting multiple players whose locations are very close to each other, which would otherwise cause one detection to block the other detection. These three channels are then combined into a single piece of data, which is then classified into the eight corresponding group labels (Safety and Cornerbacks grouped as Defensive Backs, Defensive Tackles and Defensive Ends grouped as Defensive Lineman, etc.).

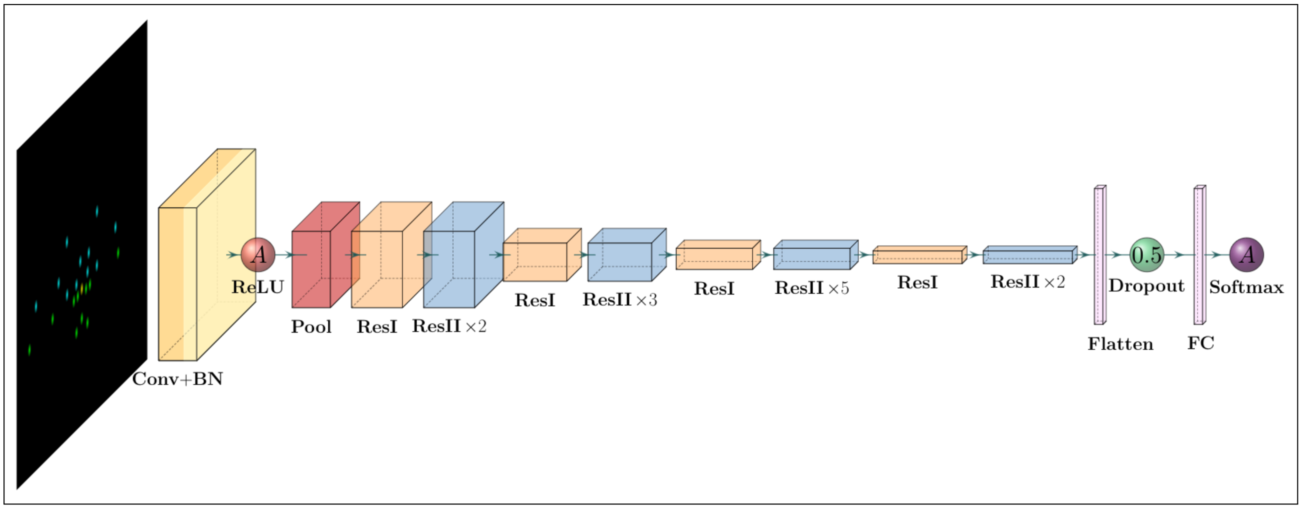


Figure 4. *ResNet Based Architecture*

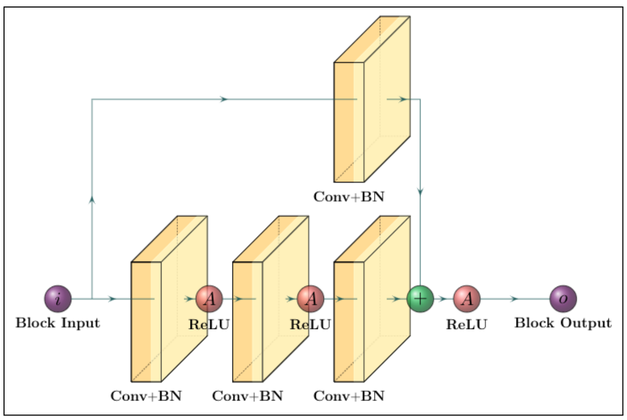


Figure 5. *ResNet Based Architecture: Block 1*

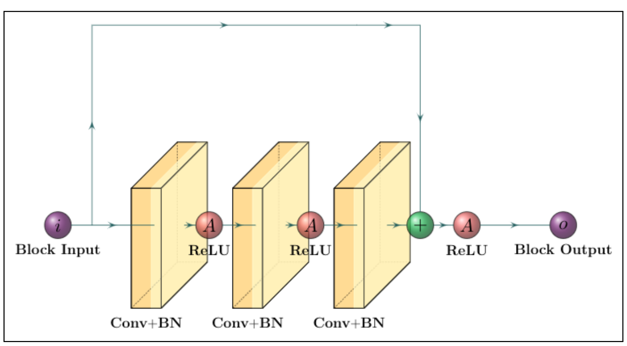


Figure 6. *ResNet Based Architecture: Block 2*

ResNet Training Results

Figure 8 shows the result from the ResNet. One problem we faced initially, as seen in Figure 8, was the misidentification of the Quarterback and the Center. They are often in the same vicinity, and the Quarterback occasionally occludes the Center. We solved this problem by forcing the output to contain a single Quarterback and a single Center. If two Quarterbacks are detected, one of their labels is changed to become the Center, and vice versa with the Center. This gives a clear improvement in the Offensive Line detections. The results of the training after this processing are shown in Figure 9.

The Precision of each of the eight groups range from 82.22% to 99.58%, while the Recall ranges from 80.43% to 99.46%. The key skill players - the Quarterback, the Running Back, the Wide Receiver, and the Defensive Back - have the following Precision-Recall values:

Precision-Recall Results of Key Skill Players

Position	Precision	Recall
Quarterback	96.67%	86.14%
Running Back	90.18%	97.12%
Wide Receiver	95.74%	95.74%
Defensive Back	99.58%	88.1%

Dividing the number of correctly labelled groups by the total number of groups gives the accuracy of the model (Equation 1). The following table shows the accuracy of individual groups, based on the data from Figure 9.

Accuracy of all Players, Offensive Key Players, and Defensive Key Players

Group	Detected Groups	Total Groups	Accuracy
Overall	1985	2199	90.27%
Offensive Key Players	482	510	94.1%
Defensive Key Players	474	538	88.1%

$$\text{Accuracy} = \frac{\#\ of\ Detected\ Groups}{\#\ of\ Total\ Groups} \qquad (1)$$

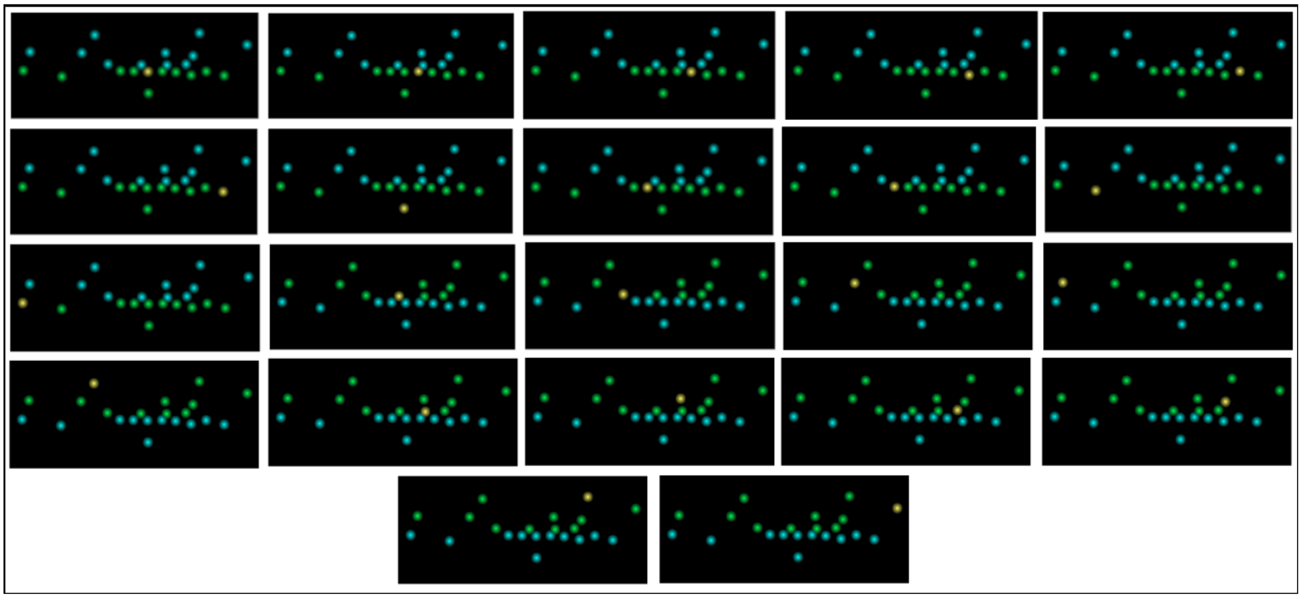


Figure 7. *Data Produced from a Single Image*

Predicted Class	Actual Class:	DB	DL	LB	OL	QB	RB	TE	WR	Total	Precision
	DB	474	0	2	0	0	0	0	0	476	0.9958
	DL	40	371	3	0	0	0	0	0	414	0.8961
	LB	24	2	183	0	0	0	0	0	209	0.8756
	OL	0	0	0	403	2	0	7	1	413	0.9758
	QB	0	0	0	90	87	3	0	0	180	0.4833
	RB	0	0	0	0	11	101	0	0	112	0.9018
	TE	0	0	0	4	0	0	74	12	90	0.8222
	WR	0	0	0	1	1	0	11	292	305	0.9574
	Total	538	373	188	498	101	104	92	305	2199	
	Recall	0.881	0.9946	0.9734	0.8092	0.8614	0.9712	0.8043	0.9574		

Figure 8. *Precision-Recall Results for ResNet Training (Pre-Processing)*

Predicted Class	Actual Class:	DB	DL	LB	OL	QB	RB	TE	WR	Total	Precision
	DB	474	0	2	0	0	0	0	0	476	0.9958
	DL	40	371	3	0	0	0	0	0	414	0.8961
	LB	24	2	183	0	0	0	0	0	209	0.8756
	OL	0	0	0	493	2	0	7	1	503	0.9801
	QB	0	0	0	0	87	3	0	0	90	0.9667
	RB	0	0	0	0	11	101	0	0	112	0.9018
	TE	0	0	0	4	0	0	74	12	90	0.8222
	WR	0	0	0	1	1	0	11	292	305	0.9574
	Total	538	373	188	498	101	104	92	305	2199	
	Recall	0.881	0.9946	0.9734	0.99	0.8614	0.9712	0.8043	0.9574		

Figure 9. *Precision-Recall Results for ResNet Training (Post-Processing)*

Future Work

We have presented a program that detects football players and labels them in groups according to their position on the field immediately before a play starts, with a focus on detecting the key skill players. This program will be foundational in future work. Our next steps will be to 1) track the players in a video feed, and 2) extrapolate useful information from this data for game analysis.

Conclusion

With recent advancements in Artificial Intelligence, automatic analysis of sports footage is rapidly becoming a popular topic. Automatic analysis of sports video with minimal human intervention is becoming possible. This automated process presents the possibility of quick sports annotations with reduced error. The research we've presented builds the foundations of further research in analyzing the game of American Football. With the results obtained from our program, we've shown the feasibility of using Artificial Intelligence to locate and label players from an overhead image of a football play before the play begins.

References

[1] NFL Football Operations, Welcome to the NFL Sideline of the Future, https://operations.nfl.com/the-game/technology/sideline-of-the-future/

[2] W. Lu, J. Ting, J. J. Little and K. P. Murphy, "Learning to Track and Identify Players from Broadcast Sports Videos," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1704-1716, July 2013, doi: 10.1109/TPAMI.2012.242.

[3] Koutsia, Alexandra & Nikos, Grammalidis & Dimitropoulos, Kosmas & Karaman, Mustafa & Goldmann, Lutz. (2007). Football player tracking from multiple views : Using a novel background segmentation algorithm and multiple hypothesis tracking. 523-526.

[4] T. Zhang, B. Ghanem and N. Ahuja, "Robust multi-object tracking via cross-domain contextual information for sports video analysis," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, 2012, pp. 985-988, doi: 10.1109/ICASSP.2012.6288050.

[5] N. Lee and K. M. Kitani, "Predicting wide receiver trajectories in American football," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016, pp. 1-9, doi: 10.1109/WACV.2016.7477732.

[6] Atmosukarto I., Ghanem B., Saadalla M., Ahuja N. (2014) Recognizing Team Formation in American Football. In: Moeslund T., Thomas G., Hilton A. (eds) Computer Vision in Sports. Advances in Computer Vision and Pattern Recognition. Springer, Cham. https://doi-org.erl.lib.byu.edu/10.1007/978-3-319-09396-3_13

[7] Ming Xu, J. Orwell and G. Jones, "Tracking football players with multiple cameras," 2004 International Conference on Image Processing, 2004. ICIP '04., Singapore, 2004, pp. 2909-2912 Vol. 5, doi: 10.1109/ICIP.2004.1421721.

[8] Hochstedler, Jeremy, and Paul T. Gagnon. "American football route identification using supervised machine learning." Proceedings of the 2017 MIT Sloan Sports Analytics Conference. 2017.

[9] Burke, Brian. "DeepQB: deep learning with player tracking to quantify quarterback decision-making & performance." 13 th MIT Sloan Sports Analytics Conference. 2019.

[10] Castin, M. (2018). Football Game Events Detection Through Features Extraction and Deep Learning. (Unpublished master's thesis). Université de Liège, Liège, Belgique. Retrieved from https://matheo.uliege.be/handle/2268.2/4536

[11] Lhoest, A. (2020). Master's Thesis : Deep learning for ball tracking in football sequences. (Unpublished master's thesis). Université de Liège, Liège, Belgique. Retrieved from https://matheo.uliege.be/handle/2268.2/9074

[12] Y. Ma, S. Feng and Y. Wang, "Fully-Convolutional Siamese Networks for Football Player Tracking," 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 2018, pp. 330-334, doi: 10.1109/ICIS.2018.8466503.

[13] Anton Meuhlemann, TrainYourOwnYOLO: Building a Custom Object Detector from Scratch, 2019, https://github.com/AntonMu/TrainYourOwnYOLO

[14] Microsoft, VoTT: Visual Object Tagging Tool: An electron app for building end to end Object Detection Models from Images and Videos, 2020, https://github.com/microsoft/VoTT

[15] JK Jung, keras-cats-dogs-tutorial, 2019, https://github.com/jkjung-avt/keras-cats-dogs-tutorial

Author Biography

Jacob Newman is currently an undergraduate student studying Computer Engineering at Brigham Young University, working in the Robotic Vision Laboratory. He will graduate in 2021, and plans to pursue a Graduate degree in Computer Engineering. His research interests include computer vision and deep learning.

Jian-Wei (Forest), Lin is currently a master's graduate student. He is studying and doing research in the Robotic Vision Lab in Computer Engineering at Brigham Young University. His work has focused on computer vision and deep learning.

Dr. Dah-Jye Lee received his Ph.D. degree in electrical engineering from Texas Tech University in 1990 and MBA degree from Shenandoah University in 1999. He served in the machine vision industry for eleven years before joining Brigham Young University faculty in 2001. Dr. Lee is a teacher, researcher, inventor, and entrepreneur. He is currently a professor and the director of the Robotic Vision Laboratory in the Electrical and Computer Engineering Department at BYU. He cofounded Smart Vision Works, Inc. in 2012 for designing and building machine vision systems with artificial intelligence for visual inspection automation. His research includes vision systems and devices with artificial intelligence, high-performance visual computing, real-time robotic vision, and visual inspection automation applications.

Jen-Jui Liu received the BS degree in industrial education in 2008, and the MS degree from the Department of Applied Electronics Technology, in 2010, both from the National Taiwan Normal University, Taiwan. Since 2020 he is a PhD student at Brigham Young University. His current research interests include Machine learning and image processing.

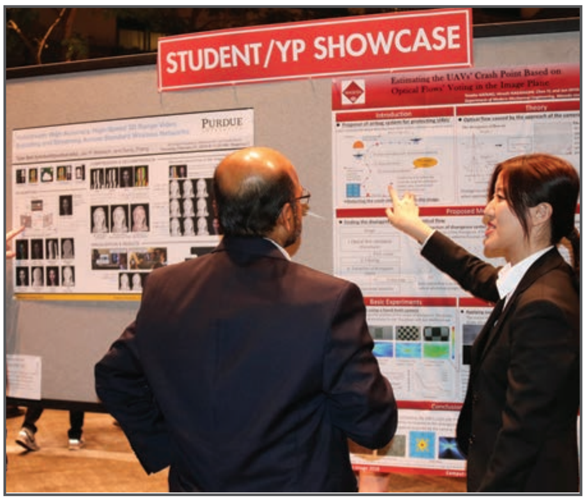
JOIN US AT THE NEXT EI!

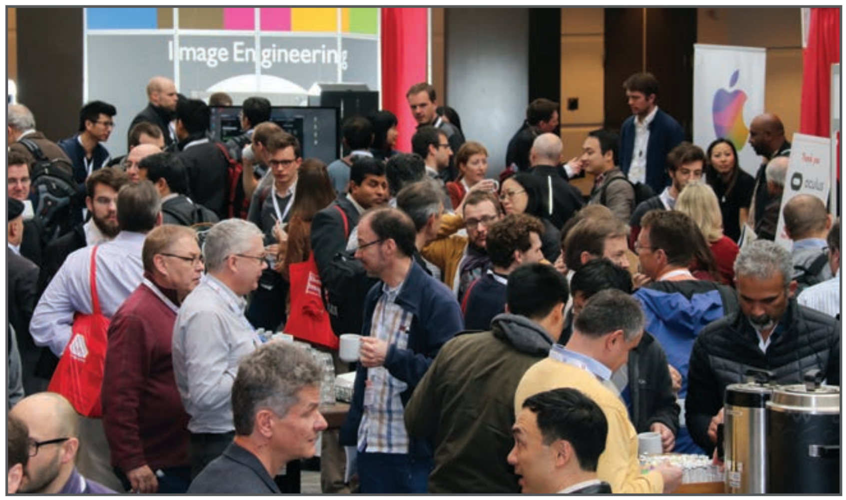
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org