**Strategic Football Play Formation Detection using Computer Vision**

by

LEONARD MARSHALL, ZAIRE WADE

ANLY 677

STATISTICAL LEARNING

*August 10, 2022*

**I. Introduction**

A significant part of game-planning for American football teams is collecting, annotating, and analyzing game videos of their own and their opponents’ games [1]. The fruit of this labor is a more competitive game-plan to effectively increase the likelihood of a winning outcome. Currently, computer vision technologies are being leveraged to extract player information from videos of football plays [1]. Several companies offer web services for facilitating this video-based game planning. For example, Pro Football Focus provides exclusive data and analysis to assists agents and agencies in landing above market rate contracts in the NFL, and to grade and assess performance of players within play designs. Deep Football is another company that provides advanced statistics, interactive dashboards, and a data platform to support interested parties. However, significant human intervention is still needed for user interfacing, organization, annotation, and video sharing. With the advent of the neural network, specifically, Convolutional Neural Networks for Computer Vision, we would like to address  the question, “How can machine vision techniques improve  video-based game planning?”

This project is the first step to automating an essential aspect of American football strategic game planning: play *formation recognition*. A formation in football refers to the position players line up in before starting a down. There are both offensive and defensive formations and many formations in both categories. Sometimes, formations are referred to as packages. It is the goal of this project to reduce the labor-intensive identification of play- formation by automating play formation recognition. If we can reduce the number of "man-hours" used to collect & annotate game film, players, coaches, and analysts will have more time to focus on resulting analytical insights.

**II. Background and Problem Statement**

         Modern sport is technology-driven and augmented, and the playing field serves as the testing grounds for cutting-edge technology. When someone tunes into a broadcast of a football game, they unconsciously absorb data and analytics in real-time. For example, a company named SportsMEDIA Technology (SMT) has partnerships with various networks. They provide producers with real-time tools to create a new standard in fan engagement. In the context of American football, they developed the "1st & Ten" line that appears on television broadcasts for NFL (National Football League) games [2]. This "yellow line" indicates the location of the first down marker to television audiences, leveraging computer vision techniques to provide an augmented reality to NFL viewers [2]. At times, a “blue line” also represents the line of scrimmage - where the players line up for the next play (Figure 1). Furthermore, SMT provides real-time object tracking technologies for NFL broadcasts, used by television broadcasters to supplement game viewing with overlaid statistics [3]. Unlike Sportsvision, the intent of this project is not to augment reality for the football broadcast viewer; it is instead meant to augment, and eventually automate, a significant aspect of game planning and analysis.

**Football Video.** American football video is organized around the concept of football plays. Each game involves a sequence of approximately 150+ plays, each lasting about 10 to 30 seconds, separated by short time intervals where no game action occurs, and the teams regroup. Videos are generally captured with a PTZ camera, showing a sideline view of the football field from an elevated location along the sideline (e.g., Figure 1). A standard video acquisition of a game involves recording a sequence of videos, one for each play in the game, which can then be automatically organized in a temporal order.

***Figure 1***

*Automated “yellow” and “blue” lines to represent the 1st down marker (line to gain) and the line of scrimmage, respectively.*

**Map

Description automatically generated**

**Play Types**. Each play has a distinct type defined by the standard play formation taxonomy in American football. This project focuses on four offensive play formation types: I- Formation, Pro-set, Single-back, and Shotgun Formation. On every offensive play, the team deploys five linemen and a QB. This is the foundation of every play. Offensive coaches then vary the number of players at the Wide Receiver (WR), Running Back (RB), and Tight End (TE) positions. This section provides a brief context of each offensive play formation. I-Formation: formation usually deploys two WRs, two RBs, and one TE. It gets its name from the two running backs lining up behind the quarterback, one behind the other, resembling an “I” dotted by the QB. Pro-set: formation uses the same personnel as the I-Formation [4]. The primary difference is how the running backs line up behind the QB [4]. The RBs line up in a split backs formation in the Pro Set, meaning one to each side of and behind the QB. Single-back: formation name implies one Running Back positioned behind the QB. The running back can line up directly behind the QB or offset either side. The role vacated by the second running back can now be deployed in several capacities making this a versatile and popular formation that usually uses between two and five WRs. Shotgun formation: The most notable feature of the Shotgun Formation is where the QB lines up. Rather than directly behind the Center, the QB positions himself about five yards behind the line, thus requiring the Center to throw him the ball [4]. The extra space gives the QB extra time to throw the ball and to give receivers additional time to run their routes.

**The Problem**. Our input will consist of individual frames of the pre-snap formations of the top plays from the years 2017-2021. Each image is categorized as either shotgun formation, or some other formation. The distinction here is useful for two reasons: first, the shotgun formation is the most distinct formation, and so an algorithm is more likely to succeed in parsing the differences between this formation and the others. However, it is meaningful as well – a shotgun approach provides the quarterback more time to prepare a throw, and thus if the shotgun formation is employed, it is more likely that a pass will take place than a running play.

**II. a) Literature Review**

This is not the first analysis to apply computer vision to football analytics and play calling. In fact, there have been several studies carried forward in this terrain in the past ten years. The approach we are taking is elementary compared to some of the research others have produced, yet it would not have been possible without their inspiration and guidance. Ojmeri and Shah carrty forward a similar analysis, leaning on classification, regression trees, naïve bayes, SVM, K Neighbors, and logistic regression in an attempt to classify play formations and find that regression trees perform best [5]. Boulier and Stekler also apply data science to football, but rather than classifying play formations, they attempt to predict the outcome of a game based on ‘power scores’ [6]. However, back in the realm of computer vision, Chen et al. also create a vision system for recognizing the sequence of plays in American football. Again, this is meant to aid coordinators and coaches in their efforts to annotate football videos [7].

Another exciting application of computer vision to the NFL involves player tracking and route identification. Made famous by Pro Football Focus, the algorithms involved in this arena aim to determine how efficient a receiver runs their route. Chu, Reyers, Thomson, and Wu apply a model-based curve clustering approach to mixed results toward this end [8]. Of course, the other side of that would relate to how efficient the defense is maintaining coverage. To this end, Dutta, Yurko, and Ventura apply unsupervised methods for identifying pass coverage among defensive backs [9]. Lee also carries forward computer vision analysis of NFL film in pursuit of valid player tracking metrics, though achieved mixed results due to challenges related to tracking players when they were not isolated [10].

There are several other research notes related to computer vision and the NFL, which can be found in the appendix of reference materials below [11][12][13][14].

**III. Data Gathering and Preliminary Methods Review**

         Throughout the course of a game, defensive coordinators for each team are tasked with assessing the intent of the opposing offense based on the formation their opposition lines up in. A defensive coordinator - and by extension, his or her defense - will be better prepared if they can anticipate the play call based on the alignment of the offense. Leveraging computer vision and machine learning to identify play formations would be just the first step to more greatly supporting defensive coordinators and their coaching staffs, but it is an important first step.

There are several serious challenges that need to be addressed as we work through this computer vision problem. First, we need to determine how best to address the variation across videos. This variation will manifest in myriad ways. We must consider the camera viewpoint, the distance that the camera is set up from the field and the action being assessed, the general quality of the footage, and the lighting and weather. So initially, there are a series of obstacles we must navigate in terms of data and image processing.

         Once we have a robust sample of clean inputs that can be analyzed by a computer, we then will be faced with the next challenge - identifying and distinguishing offensive players from defensive players, from referees, and from any other interfering elements in a frame. This exercise of cleaning the data and images is of utmost importance, as the quality of the analysis will only be as good as the quality of the inputs to the models used.

         There are various likelihoods of play-calls based on the personnel, packages, and play formations that are present on a field at a given time. For example, certain formations are more likely to lead to a running play or a passing play. Understanding the alignment (along with the tendencies of the personnel on the opposing team) can influence the outcome. If a defensive coordinator can quickly understand the formation that an opposing offense is presenting, they are more likely to be able to align their defensive players to stop the opponent on that play.

**III. Method & Design**

Player tracking data has been essential to the advancement of American football analysis since the National Football League (NFL) released its Next Gen Stats tracking data publicly for the first time in December 2018. This project aims to build upon prior American football player tracking methods using computer vision techniques. While tracking datasets in other sports often contain detailed annotations of on-field events, annotations in the NFL's tracking data are limited. In addition, creating these annotations typically requires extensive human labeling, which is difficult and expensive.

We tackle this class of problems by creating annotations for offensive formation types by player's location using unsupervised learning techniques, which require no manual labeling or human oversight. We define a set of features from the NFL's tracking data that help distinguish between offensive formations—using mixture models to create clusters corresponding to each offense group, allowing us to provide probabilistic assignments of each formation type (or cluster). We evaluate the performance of our detection and classification framework by applying it to real-world football videos. No publicly available dataset accurately represents video captured and used by American Football teams. Typically, a football coach would segment the footage of a whole game into single play clips, identify their respective formation frames, and then classify the formation type.

The dataset we seek to compile will comprise of many distinct offensive plays. The football game videos will be collected from both NFL and college games showing a sideline view of the game. The videos will then be manually segmented such that each clip shows a single play. In our dataset, we seek only to analyze play clips that are the standard resolution (640 × 480 resolution) or high definition (1440 × 1080 resolution). The formation frame, line of scrimmage, and formation label serve as our imagery's ground truth data.

Our proposed solution aims to identify the formation of the team playing offense in a football game. The input to our framework is a single football video play clip. The proposed approach consists of two major modules: pre-processing the play video clip and formation recognition of the offensive team in the play footage. The ‘empirical design’ and ‘analytical steps’ sections will detail these approaches in more detail. However, the second phase of formation recognition will have the greatest pay-off for the analysis. This phase of analysis will determine the success of the work. Given a registered play clip, the framework will first identify the frame in which the two teams are lined up in formation. Once this frame is detected, the field line that separates the two teams at formation time (otherwise known as the line of scrimmage) is determined in the formation frame. Next, we utilize the spatial distribution of the two teams to identify the offensive team. Finally, we extract features from the offense region to train a linear SVM to classify the different offensive formations.

**III. Empirical Design**

Because the goal of the analysis is to detect and identify formations within still frames of offensive formations, the first stage of this process includes collating these images. We will scour video sources such as youtube.com to compile these images across several professional and amateur games. In other analyses in this research field, researchers have had access to game film via formal NFL channels. However, due to the fact that this analysis is being carried forward in the summertime, which is the off-season for the NFL season, subscription to these services is currently closed.

The data will be collected from game footage from September, 2015 through February 2021. In addition to the images, we will also compile a dataset with 1:1 relation to these images. This dataframe will represent the ‘ground truth’ of the formation. The classification of these formations will result in a dataset with four categories: ‘shotgun’, ‘single-back’, ‘i-formation’, and ‘pro-set’. The human eye can detect and identify these formations, but the overall intent of this analysis is to automate this process. In the case that a coordinator or interested party intends to analyze hundreds of formations and pre-play sets to prepare for a season ahead, this categorization exercise would be exhausting for a human.

After the images are uploaded into a dataframe in a python environment, we will next need to process them. To do so, we intend to first identify the line of scrimmage. Typically, the camera angle will not be aligned perfectly along the line of scrimmage and so the first stage to correct for this (so that we may separate the defensive from the offensive players) will be to rotate the image such that the line of scrimmage runs vertically. With this accomplished, the defensive players are actually superfluous to our analysis. In all likelihood, we will crop the images to remove the defensive alignments altogether.

There still may remain a few more issues to work out in this detection process. For one, it is possible that the referee may be in the frame. Another challenge is the fact that there are some players whose positions will only change marginally from formation to formation. In the case of this analysis, where they line up will represent ‘noise’, so it may behoove us to crop the images even further. There will be a few ‘zones’ that we are most interested in as we take the analysis forward.

The primary zone of interest is the backfield - essentially, the area behind the offensive line. The quarterback always lines up behind the center. For ‘i-formation’, ‘pro-set’, and ‘single-back’ alignment, the quarterback lines up directly under the center. But for ‘shotgun’ alignment, the quarterback is further back. Our null hypothesis is that the ‘shotgun’ alignment will therefore be easiest to detect and identify.

As we crop and specify the areas of interest in each photo, and properly rotate the images such that they are consistent from one to the next, the following challenge will be to detect the players in their given locations. To do so, we will rely on various thresholding techniques. Using the color of the jerseys as a primary signal, we will hope to detect these players in the field. In some cases, the playing field will appear green, as most fields are grass, while a few others are artificial turf. However, there may also be some images in our dataset that are played on snow. In either case, the color of the players’ jerseys will be different than that of the playing field. With our prior knowledge of the teams we are analyzing, we will have a strong idea of which colors we are trying to isolate. The independent variables we need to identify the formations are largely driven by the coordinates of the players within the image relative to one another. As it relates to data pre-processing, figure 2 details the initial image (figure 2) and the final frame for analysis (figure 3).

**Figure 2**

***Initial image prior to processing.***

A football game in progress

Description automatically generated with medium confidence

**Figure 3**

***Final image for entry to linear SVM analysis model. Cropped to remove some defenders, rotated so the line of scrimmage runs vertically.***

A group of people on a football field

Description automatically generated with low confidence

**III. Discussion/Results**

At a high level, the results of our analyses are still pending. There are several issues that are yet to be resolved, and in pursuit of further publication we will hope to revisit these analyses. Many of these methods are being learned as they are applied, and thus more time is required for truly robust findings.

An initial challenge for this research was finding images to analyze. Other research, as mentioned, analyzed this data terrain as well. But in those instances, many of the researchers were able to leverage NFL-All 22 – a subscription video service that is available during the regular season. Due to this analysis being carried forward during the off-season, we instead had to resort to Youtube to source our images. While these images were consistent in their dimensions, they were not consistent in nearly any other way. The angle of the camera toward the action varies (both the altitude above the field, and the direction the offense is moving on the field). Some broadcasts employed ‘loud’ graphics such as the dark blue, Line of Scrimmage artificially embossed on the field of play, while others did not. The color of the jerseys varied significantly from image to image. The field conditions (e.g., snowing, raining), also impact the quality of the images and our ability to apply consistent methodology in pursuit of cogent results. Another challenge manifested due to simple hardware constraints. With no access to cuda, we rely on cpu power rather than gpu power, and this handcuffed our ability to process more than the 404 images in a timely manner, or to attempt more than ~5 epochs during various unsupervised method applications.

Nevertheless, we did make attempts to elevate our code above and beyond the simple lab work (though, attempts were mostly futile!). Initially, there is the stage of image processing and transformations. As mentioned, the characteristics of the images varied significantly from one image to the next. While our goal, ultimately, is to distinguish whether or not a quarterback is lined up directly behind the center or several yards behind, it proved a challenge to consistently identify the same region from one image to the next in order to hone in on the region of interest.

Where we did find (resounding) success was in the domain of image classification with VGG convolutional neural networks. VGG models are a type of CNN Architecture proposed by Karen Simonyan & Andrew Zisserman of Visual Geometry Group (VGG), Oxford University, which brought remarkable results for the ImageNet Challenge. Using these pretrained models, we achieve 100% prediction accuracy in both the training set and testing set, and the residual train and test loss across the epochs is also rather low, and diminishes quickly (see figure 4).

**Figure 4**

***The train and test losses graphed against the epochs. We see that by the third epoch, the train loss regresses to nearly zero, while the test loss gradually declines across the 5 epochs.***

Diagram

Description automatically generated

Ultimately, we remain suspicious of the great level of success the predictive powers of the VGG CNN application. It is shocking that while other methods struggle or fail, the VGG application predicts all shotgun and all under-center formations flawlessly, with a test and training accuracy of 100%.

This stands in direct contrast of our work involving InceptionV3 Evaluation. In this application of CNN regardless of all attempts to finetune the approach, we find that the algorithm predicts all formations to be of the shotgun variation, even though just ~70% of the formations are shotgun. This is not an issue we were able to resolve, and so between the VGG model being perfectly perfect, and the InceptionV3 method being perfectly imperfect, it is clear that more time needs to be spent finetuning these findings before advancing in the research process.

**V. (Brief) Conclusion**

Ultimately, there were myriad issues and challenges that arose throughout the duration of this project. However, it was still a really exciting opportunity to get comfortable with image transformation approaches, image classification algorithms, and applications of neural networks and deep learning to a really complicated terrain in image mining and computer vision. It would be a worthy endeavor to retrace the steps of this project at a future time – when better footage and images are made available during the NFL regular season. When it comes to a project in the space of image mining, your outputs are only ever going to be as good as your inputs. Our team was significantly hamstrung by our lower quality images. To circle back with better images and a more regimented data processing approach, paired with our already skeptical and curious constitution would elevate the findings of this research significantly. Nonetheless, the value of the ambition of this research is clear.

**References**

[1] S. Chen et al., "Play type recognition in real-world football video," IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 652-659, doi: 10.1109/WACV.2014.6836040.

[2] Lee, T. (n.d.). *Automating NFL Film Study: Using Computer Vision to Analyze All-22 NFL Film*. Retrieved from https://cvgl.stanford.edu/teaching/cs231a\_winter1415/prev/projects/LeeTimothy.pdf.

[3] *Description of Football Technologies*. Sportvision . (n.d.). Retrieved from www.sportvision.com/football

[4] *Football formations for beginners*. PRO TIPS by DICK'S Sporting Goods. (2019, May 1). Retrieved June 3, 2022, from<https://protips.dickssportinggoods.com/sports-and-activities/football/football-101-football-formations>

[5] Omar Ajmeri and Ali Shah. Using Computer Vision and Machine Learning to Automatically Classify NFL Game Film and Develop a Player Tracking System. page 9, 2018.

[6] Bryan L. Boulier and H.O. Stekler. Predicting the outcomes of National Football League games. International Journal of Forecasting, 19(2):257–270, April 2003.

[7] Sheng Chen, Zhongyuan Feng, Qingkai Lu, Behrooz Mahasseni, Trevor Fiez, Alan Fern, and Sinisa Todorovic. Play type recognition in real-world football video. In IEEE Winter Conference on Applications of Computer Vision, pages 652–659, Steamboat Springs, CO, USA, March 2014. IEEE.

[8] Dani Chu, Matthew Reyers, Lucas Wu, and James Thomson. NFL Big Data Bowl.

[9] Rishav Dutta, Ronald Yurko, and Samuel Ventura. Unsupervised Methods for Identifying Pass Coverage Among Defensive Backs with NFL Player Tracking Data. arXiv:1906.11373 [stat], June 2019. arXiv: 1906.11373.

[10] Timothy Lee. Automating NFL Film Study: Using Computer Vision to Analyze All-22 NFL Film. Class Project, Stanford University CS231A.

[11] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs], April 2015. arXiv: 1409.1556.

[12] Shiladitya Sinha, Chris Dyer, Kevin Gimpel, and Noah A. Smith. Predicting the NFL using Twitter. arXiv:1310.6998 [physics, stat], October 2013. arXiv: 1310.6998.

[13] Nathan Sterken. RouteNet: a convolutional neural network for classifying routes. page 12.

[14] Ronald Yurko, Francesca Matano, Lee F. Richardson, Nicholas Granered, Taylor Pospisil, Konstantinos Pelechrinis, and Samuel L. Ventura. Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. Journal of Quantitative Analysis in Sports, 0(0), January 2020.