NFLFormation: A Framework for Offense Formation Recognition in American Football

Hadi Ismail
Radu-Mihai Savancea
Tobias Schmid

University of Stuttgart

October 28, 2025

Contents

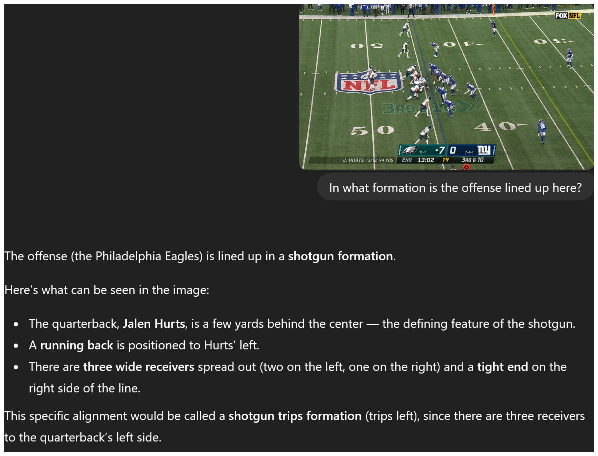

Figure 1: Chat-GPT is shown a balanced shotgun formation

Abstract

We investigate the ability of foundation models to perform the task of formation recognition in American football. Three approaches are compared. The capabilities of a baseline foundation model to guess the offense formation based on a pre-snap image is investigated. We then observe if performance improves by providing player information in the prompt, to better understand potential flaws in model reasoning. Lastly, we propose an automated approach in three steps, using an image segmentation model, as well as a lightweight model fine-tuned on data specifically annotated for player and formation classification.

1 Introduction

Because of its unique play structure American football poses interesting tactical considerations based on pre-snap information. Within the rules of the game, the offense is able to set up before each play in a wide variety of formations. From a defensive standpoint, the game plan to stop an offense often results from pre-snap information (that is, the information a player/coach is able to gather between plays, before the ball is brought into play), based on offensive formation. For example, defenses often determine the number of defenders responsible for stopping the run (so-called *box defenders*) based on the number of offensive players in the box (the space around the offensive line). The task of identifying the formation the offense is in is called formation recognition, and is the main problem tackled in this work.

Through platforms like Hudl, even amateur teams are able to film their practices and games for review and analysis. A key aspect of such an analysis is the identification of tendencies based on formation. In order to perform such an analysis, coaches need to enter information about each formation pre-snap by hand. In addition to being rather time consuming, bad video quality and angles, as well as a lack of unified language, often increase task complexity.

Foundation models might be able to support in the automation of data entry tasks in offense formation recognition. However, broadly trained models are susceptible to hallucinations, as can be seen in the example below.

As demonstrated in 1, models like Chat-GPT are able to identify a lot of information relevant to the offensive formation on their own, even without training. When asked to identify the offense formation in 1, Chat-GPT provides a semi-correct answer.

Correct information:

- The model is able to identify the offense (team in possession of the ball) as the Philadelphia Eagles
- The quarterback is identified by name (possibly because of the written information provided in the picture?)

- The quarterback is correctly identified to be in a shotgun alignment, about 5 yards behind the center
- The running back aligns to the left of Jalen Hurts
- The model furthermore recognizes three wide receivers (two on the left, one on the right)

Incorrect information

- ChatGPT identifies a "Tight End" on the right side of the line. While the type of player lining up in this picture is a Tight End on the roster report, the position a few yards away from the line, as well as his stance, classify him as a receiver (a Split End to be exact) on this play
- The name of the alignment as "trips left" is incorrect. While the running back is technically an eligible receiver, a trips alignment has three receivers to one side, and one receiver to the other, with the running back in the box

We can observe that the model is able to identify key information in the picture. However, the conclusion the model reaches, that the offense is aligned in a trips formation, is false. An important question is whether the model makes this mistake based on a misinterpretation of the picture, or a lack of understanding of offensive formations. To understand the behavior of the model better, we want to investigate if the model performance can be improved by providing player information in the prompt. Furthermore we want to investigate a two-step formation recognition approach with player identification and formation classification.

2 Related Work

Being a billion-dollar industry in North America, the NFL and other major American football leagues are of great interest to the media, betting firms, and general audiences. A well-known challenge in this field is the NFL Big Data Bowl, a Kaggle competition that focuses on tabular data and aims to predict, from various player and play features, the number of yards a player will gain or the probability that a defender makes a tackle, based on factors such as location, angle, and speed. Beyond tabular challenges, at the time of writing this project proposal, we are not aware of any research papers that use foundation models to address this task. Other vision-related projects use data collected from video games such as Madden to perform player labeling using classical CNN architectures [NSTL23]. More recent papers, such as [HOPL24], identify a limited number of player formations using MLP-based architectures trained on real-world game data samples. In [OPL24], an LSTM-based architecture was used to determine whether a frame was captured pre-snap or post-snap, utilizing player detection methods beforehand.

3 Proposed Methods

The task of formation recognition can be broken down into three sub-tasks:

1. **Player identification**: An image segmentation task. Given an image, return all the areas, that contain a (offense) player
2. **Player classification**: A classification task. Given an image with identified players, return the position and alignment of each player (on offense)
3. **Formation classification**: A classification task. Given an image, as well as a list of offense players, with positions and alignments, return the name of the offense formation

We want to compare three approaches. The first approach, the **baseline**, treats the task of formation recognition as an end-to-end task. We feed a foundation model like Chat-GPT with a picture of the pre-snap alignment and ask the model to classify the formation by returning a formation name.

In the second approach, called the **oracle** approach uses the same model as the baseline. However, the model will also receive gold-standard player information (alignment and position of each player) as well as the pre-snap image. We want to see if model performance improves by expanding the prompt, compared to the baseline model. If this is the case, errors of the baseline model might be the result of poor image understanding, compared to a lack of knowledge on offense football formations and / or terminology.

Finally, we propose an automated two-tiered **fine-tuned** approach. First, we use a foundation vision model like SAM to detect players on the field and then fine-tune a lightweight pre-trained model on both player classification (input = segmented picture, output = list of positions and alignments) and formation classification (input = list of positions and alignments output= formation name).

4 Dataset

For the purpose of fine-tuning, we need a labeled dataset, suitable for the tasks of formation recognition. We use the publicly available NFL Jersey Tracker dataset [Tra25] from Roboflow as a starting point. This dataset already contains a number of pre-snap images with player identification. To increase the number of data points, we add self-created samples using SAM [KMR+23] as a segmentation foundation model. For both the public data, as well as our own, we will manually annotate player information (position and alignment) for the player classification task, as well as the formation name for the formation classification task.

5 Project Goals

The goals of this project are tied to the three model approaches. Using the baseline approach, we want to investigate the capabilities of general foundation models for offense formation detection. For the oracle approach, we want to compare formation recognition as an end-to-end task, versus a task split into player identification and classification, and formation classification. That way, we want to understand if model performance issues stem from picture interpretation or a lack of understanding of American Football theory, and if we can improve performance by providing additional information in the prompt. Most importantly, with the last approach we investigate the feasibility of an automated framework that can correctly classify offense formations, possibly outperforming larger, more general models.

References

[HOPL24] A. Hong, B. Orr, E. Pan, and D.-J. Lee. American football play type and player position recognition. *Electronics*, 13(18), 2024.

[KMR+23] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[NSTL23] Jacob Newman, Andrew Sumsion, Shad Torrie, and Dah-Jye Lee. Automated pre-play analysis of american football formations using deep learning. *Electronics*, 12(3), 2023.

[OPL24] Benjamin Orr, Eric Pan, and Dong-Jun Lee. Optimizing football formation analysis via lstm-based event detection. *Electronics*, 13(20):4105, 2024.

[Tra25] Football Tracking. Football jersey tracker dataset. `https://universe.roboflow.com/football-tracking/football-jersey-tracker`, jul 2025. Accessed 2025-10-27.