

Methodology of the Formation of Sports Matches Statistical Information Using Neural Networks

Olena Sorokivska¹, Iaroslav Lytvynenko¹, Oleksandr Sorokivskiy², Halyna Kozbur¹, Iryna Strutynska¹

¹ Ternopil Ivan Pulu National Technical University, 56, Ruska Street, Ternopil, 46001, Ukraine

² IT-Company Amazinum, Ternopil, 46001, Ukraine

Abstract

The article develops a methodology for the formation and processing of statistical information about held sports matches using neural networks. To find the players and the ball in the video, the authors used the Yolov5 Model. It is fast and accurate, being ideal for such a task. To determine the position of the players on the pitch, pix2pix neural networks were used. They were trained on images of players and their positions on the pitch, allowing them to accurately identify players' positions in new images.

The technology called soccerreid was used to re-identify players. It allows to distinguish one player from another based on their appearance and movements. Also, the creation of an API in Python using Flask to obtain statistical information about players and their actions on the pitch was described. This will allow coaches and analysts to receive valuable information to improve team strategy and tactics.

The elaborated automated method of generating statistical data has great application potential in the industry of football match analysis. The results can be used for further research in the area of automating of the statistics formation. The obtained results can also be used to improve work in other areas related to sports. It is also worth noting that the use of artificial intelligence technologies, which are used in the development of such an intellectualized methodology, can significantly facilitate and speed up the process of analyzing video recordings, which will increase the efficiency of work and reduce the cost of time and money.

Keywords

Machine learning, deep neural networks, football, computer vision, homography, YOLO.

1. Introduction

The development of the methodology for the formation of statistical information obtained on the basis of the analysis of video recordings of football matches is a very relevant topic of research, especially in the context of modern technologies and the popularity of football in the world. It is possible, with the help of such a technique, to provide an accurate and objective analysis of the game, identify the strengths and weaknesses of the teams, and also help the coaches in preparing for the next matches. In addition, the collection and analysis of statistical data from video recordings can be useful for organizing competitions and improving their level, increasing the interest of spectators in the game, as well as the development of the football industry as a whole. It is also worth noting that the use of artificial intelligence technologies, which are used in the development of such an intellectualized methodology, can significantly facilitate and speed up the process of analyzing video recordings, which will increase the efficiency of work and reduce the cost of time and money.

Proceedings ITTAP'2023: 3rd International Workshop on Information Technologies: Theoretical and Applied Problems, November 22–24, 2023, Ternopil, Ukraine, Opole, Poland

EMAIL: soroka220996@gmail.com (A. 1); iaroslav.lytvynenko@gmail.com (A. 2); Gosasha401@gmail.com (A. 3); kozbur.galina@gmail.com (A. 4); strutynskairy@gmail.com (A. 5).

ORCID: 0000-0001-8549-2910 (A. 1); 0000-0001-7311-4103 (A. 2); 0009-0006-6477-5878 (A. 3); 0000-0003-32970776-2910 (A. 4); 0000-0001-5667-6569 (A. 5).



© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

The purpose of the article is to develop a methodology for the formation of statistical data of personal indicators of players based on the processing of video recordings of past football matches. The main goals of the study are to provide an efficient and automated process of obtaining statistics about players based on video recordings of matches.

The developed automated method of generating statistics from the recordings of football matches makes it possible to optimize and improve the existing processes in the coach's work, reduce the number of errors and accelerate the development of the team. The results can be used for further research in the field of automating the formation of statistics. The obtained results can also be used to improve work in other areas related to sports.

2. Related Works

In an early work of Assfalg, J., Bertini, M., Colombo, C., Del Bimbo, A., and Nunziati, W. [1] a method of finding a homography based on the position of the pitch and correspondence of lines was considered. The disadvantages of this technique are that its aim is only the gates and corner parts of the pitch. All work is based on the annotation of events, namely a goal kick, a corner kick and a free kick from the goalkeeper. Also, the work developed an algorithm for determining the type of shot based on the location of the players on the pitch. Finding is done by highlighting them by color. Classification of events is based on the placement of players on the pitch.

If we consider works based on the study of useful statistics for teams, one of the most popular is the work of Perin C., Vuillemot R. and Fekete J. [2], which describes the importance and methods of qualitative analysis in football. Scientists suggest using a special visualization for corner kicks, long runs, pass clusters, shot distribution and others. The statistics and visualization in the work are explored in depth, but the drawback is that it lacks information about the players and their position on the pitch. This information is only available for labeled data.

A recent paper by Stein, M., Janetzko, H., Lamprecht, A., Breikreutz, T., Zimmermann, P., Goldlucke, B. and Keim, D.A. [3] deals with the full cycle of obtaining and visualizing statistical data from video recordings. To find players on the pitch, the authors use the segmentation method based on colors. To determine the position of players on the pitch, the SIFT method [4] is used, the essence of which is to find key points between images, obtaining vectors from them for combining images. The work states that the first two minutes of the match are enough to get all the camera angles and connect the images together. This allows you to get a complete diagram of the pitch. Once the schema is obtained, a predefined homography matrix is used to obtain the changes for the next frame. In general, the accuracy of these methods is not specified. In addition, the authors themselves point out that the methods of finding players on the pitch are not accurate enough. Also, the system does not include the function of automatically finding the ball on the pitch, which limits the calculation of statistical information in the future.

In modern research S. Afzal, S. Ghani, M. M. Hittawe, S. F. Rashid [5] explore the current state of the art at the intersection of visualization and visual analytics, as well as image and video data analysis. The authors classify visualization articles based on various taxonomies used in visualization and visual analytics research. They review these articles in terms of task requirements, tools, datasets, and application areas, and discuss ideas based on the results of their survey. Basic scientific research is being conducted by modern scientists to evaluate the control zone in badminton doubles games using information from drones [6], as well as supplementing basketball videos with built-in gaze-guided visualizations [7]. M. B. Jurca [8] in his work constructed a robust pipeline for the sports analysis community in order to successfully extract useful information from broadcast football matches. The scientist proposed a fast and efficient solution, based on computer vision and machine learning methods and algorithms. Scientists S. Rahimi, A. Moore, P.A. Whigham [9], based on the agent's conceptual space-time model and reasoning behavior, developed guidelines for the design of a realizable vector-agent model. They applied sensitivity-variability analysis to measure the performance of different configurations of system components with respect to new movement patterns. Researchers Z. Chen, J. Beyer, H. Pfister, Q. Yang, H. Xia, X. Xie, Y. Wu [10], focused their attention on supplementing sports videos with natural language. However, none of the mentioned articles gives a clear and sufficiently

comprehensive answer to the question of how to ensure an effective automated process of obtaining statistical data about players based on video recordings of matches.

3. Proposed Methodology

One of the subtasks in the methodology of statistics formation is determining the position of players on the pitch. At the same time, the main information is the frames from the video. Solving this subproblem is impossible without taking into account additional variables. We offer several basic methods of solving this problem.

3.1 Determination of the homography matrix by manual method

This method involves manual adjustment of one or more cameras. In order to determine the position of the players on the pitch, you need to know the boundaries of the pitch, the distance from the base of the camera to the pitch, the distance from the top of the camera to the pitch, and the homography matrix. A homography matrix is a transformation matrix used to represent three-dimensional objects in two-dimensional space or to perform other similar tasks in graphics processing. Homography is a mathematical concept that describes the relationship between two projective spaces that can be realized as images of three-dimensional objects on a plane.

The homography matrix is usually determined by calibrating the camera and fitting to the points that represent the image on the plane. Knowing the homography matrix, one can perform image transformation operations such as scaling, rotation, and translation. In a more general sense, the homography matrix is used to describe transformations between spaces of any number of dimensions, and is not limited to the field of graphics processing. Therefore, in this approach, the matrix is adjusted manually. First, one wide-format camera or several cameras are placed in a fixed position, key points are selected from 3D space and 2D space, and the homography matrix is found.

In the case of one camera, it needs to be set evenly, because an error of a few millimeters can become an error of several meters, due to the fact that the football pitch is large - about 100 meters. After finding the homography matrix for the initial position of one camera, it is necessary to update the matrix using information about camera rotation angles and zoom.

3.2 Determination of the homography matrix by an automatic method

The automatic method is based on the determination of camera parameters based on the gradient analysis of image edges. The normal camera calibration process must use points on the image that can be difficult to determine and may require additional training. At the same time, the edges of the image, which can be easily determined, have a high sensitivity to changes in camera parameters, which made them an object of research.

The authors propose to use the gradient of the edges of the image to define the parameters of the camera without the need to use points. The method for determining the camera parameters consists in finding the optimal mapping between the corresponding edges in two images, using the gradients of these edges. The research was conducted on different types of images and it was demonstrated that the method can be effective for determining camera parameters, especially in situations where the points in the image are difficult to define, but the edges of the image can be easily found.

The main result of the study is the use of image edge gradients to determine camera parameters, which may be useful for developing more efficient and accurate camera calibration methods in the future. An example of the result of this approach is shown in Figure 1.

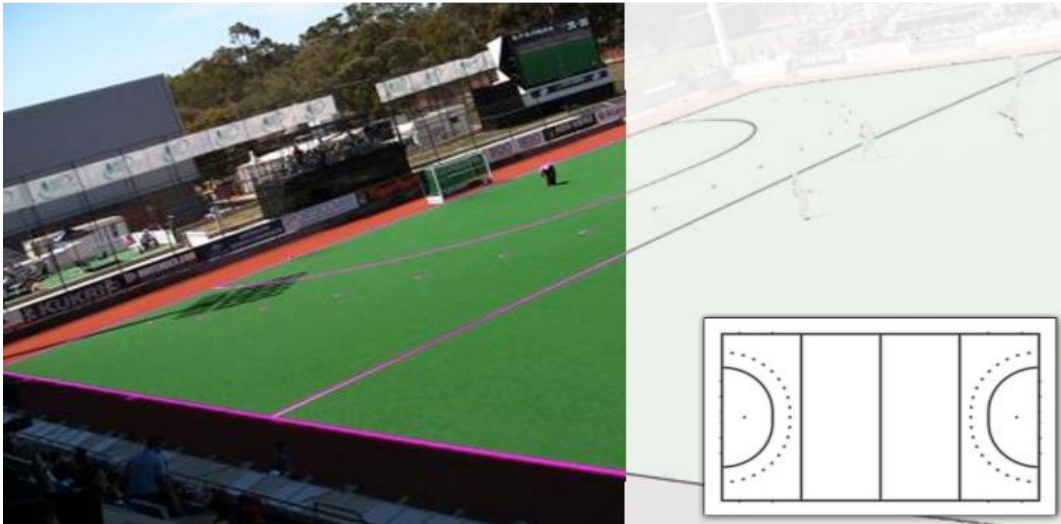


Figure 1: An example of the result of the gradient approach

Another approach to automatically determine the homography matrix is to develop a neural network that determines the positions of key points in the image to find the homography. For this, a similar model architecture to UNet is used, using an encoder and a decoder. U-Net is a deep neural network architectural approach for semantic image segmentation, that is, for dividing an image into subregions and assigning a class to each subregion. It is widely used in biomedical imaging, in particular in the tasks of segmentation of cells, tissues, organs and pathological changes.

Net architecture consists of two main parts: an encoder and a decoder. The encoder consists of several convolutional layers (convolutional layers) and pooling layers (pooling layers), which reduce the size of the image and increase the number of channels. This part of the network performs the function of feature extraction, which is then used by the decoder.

The decoder consists of transposed convolutional layers (transposed convolutional layers) and concatenation with previous encoder outputs. The decoder gradually increases the image size and decreases the number of channels to obtain a segmentation map. Concatenation helps transfer local information from the encoder to the decoder, allowing for a more detailed segmentation map. In general, the encoder provides the extraction of a hierarchy of features that can be used for image segmentation, and the decoder provides the reproduction of the image from the segmentation map using the information from the encoder.

Next, the training images were labeled using 91 key points. Also, the usual calculation of weights for filters in the convolutional model has been replaced by dynamic filter generation. For this, each of the 91 points was recoded into a vector and was used to train the decoder model.

IoU metric was used to measure accuracy. Using World Cup Dataset with the help of the algorithm an accuracy of 98% was achieved. An example of the result of the algorithm is shown in Figure 2.

Another approach is to determine the homography matrix using synthetic data. This approach uses a dual GAN model to obtain a picture with pitch corners. GAN (Generative Adversarial Networks) are neural network models that consist of two deep neural networks: a Generator and a Discriminator. The generator creates new images while the discriminator tries to distinguish them from the real ones. The model is trained in a confrontation process between the generator and the discriminator, where the generator tries to create such images that cannot be distinguished from real ones, and the discriminator tries to distinguish between real and synthesized images. In the learning process, the generator and the discriminator interact and learn from each other, improving their quality and ability to produce realistic images.

The generator and the discriminator are competing models that work together to provide GAN training. The generator takes a random noise or vector as input that is used to generate new images. The discriminator takes an image as input and determines whether it is real or generated. The most important element of GAN is the loss function, which measures the error rate of the generator and the discriminator. The loss function must be configured so that the discriminator can distinguish between real and generated images and that the generator produces images that are close to the real ones. In the

process of GAN training, the generator tries to change its parameters to reduce the discriminator error and improve the quality of its images.



Figure 2: An example of the result of the algorithm using a neural network

After obtaining information from the GAN model, HOG transformations are used to obtain information about the image, which then selects the best match from a database containing similar images and their corresponding homography matrices. HOG (Histogram of Oriented Gradients) is a method for determining features in images, which is widely used in computer vision and image processing. HOG is based on the concept that the shape of objects can be determined based on the orientations of pixel gradients in an image. The resulting vectors are considered image features that can be used to recognize objects in the image. An example of the algorithm is shown in Figure 3.

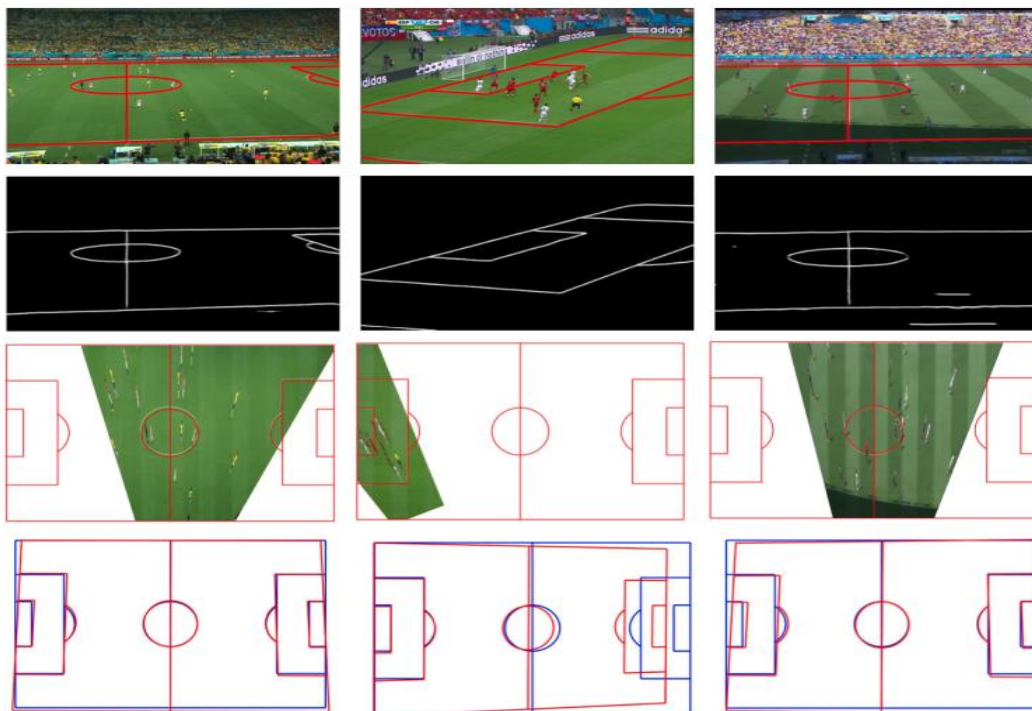


Figure 3: Example of homography definition using GAN models

However, the image shows that the model does not always predict the position of the pitch as accurately as possible. The maximum accuracy of this task is beyond the scope of work.

3.3 Analysis of approaches to player re-identification

Re-identification of players on the pitch is an important task for analyzing football matches and understanding how players move and interact on the pitch. Computer vision can be used to automatically identify players based on their appearance.

A system called Torchreid is commonly used to re-identify people. Torchreid is an open source software (open-source software), which provides a framework for developing and experimenting with image-based re-identification algorithms. This framework provides support for numerous datasets,

including Market1501, DukeMTMC-reID, CUHK03, and others. In addition, it contains tools for data processing, model building, result validation, and visualization.

Torchreid main features:

1. Support for various neural network architectures for re-identification, including ResNet, DenseNet, NASNet, EfficientNet, etc.
2. Support for various loss functions, including cross-entropy loss, triplet loss, quadruplet loss, circle loss etc.
3. Support for various methods to reduce data dimensionality, including Principal Component Analysis (PCA), Linear Discriminator Analysis (LDA), t-SNE, etc.
4. Ability to build hybrid models that use images and video simultaneously to improve re-identification results.
5. Support for various methods for data collection, including manual annotation, automatic annotation using computer vision algorithms, etc.
6. Ability to use pre-trained models to achieve better results on new data.
7. Support for various metrics for evaluating re-identification results, including Cumulative Matching Characteristics (CMC) curve and mean average precision (mAP).

Re-identification technology SportsReID is based on this technology for finding football players, which is a system for re-identification of players in video matches using computer vision. The system uses deep learning technologies, including neural networks, to determine the identity of players in videos. To do this, the video is first processed, during which frames with the image of the players are extracted from it and algorithms are used to determine features such as body shape, clothes, shoes and other details of the appearance.

The obtained features are compared with data from a database containing images of players and their identification data. The search is performed using image comparison algorithms such as histogram oriented gradients (HOG), deep neural networks, and others. SportsReID is used in large team sports events where it is necessary to track the movements of many players at the same time. The system can help in training teams, analyze the game and identify the best players in order to improve performance.

4. Results

4.1. Highlighting the main requirements for the methodology of the formation of statistical information about the held sports matches on the basis of video recordings

In the process of the research, the authors started from the task that the method for generating statistical information should be able to find players and the ball in video, determine their position on the pitch and calculate variable data that can be used by coaches in the future to improve the team's work.

The methodology should also be able to be integrated into any application. For the successful integration of the technique into applications, it was chosen to create an API for accessing the technique, which can be used to obtain information about the movement of players and calculated statistical data.

Statistical data that should be in this technique for each player:

- Time of ball possession;
- Distance traveled;
- Number of passes;
- Number of interceptions;
- Time on your side;
- Time spent on the side of the competitor;
- Average speed.

Due to these statistics, you can make a complete picture of the player's work during the game, as well as understand the key advantages and disadvantages of certain strategies.

4.2 Finding the players and the ball

A dataset with placed players and a ball was used to train ball detection. The players and the ball were highlighted by rectangles, and the coordinates of the rectangles were recorded in a special format for training the model. As a result, 1000 pictures were selected for training the model, which contained, on average, 14.5 players in the frame. An example dataset is shown in Figure 4.



Figure 4: An example of a dataset for finding players and the ball

In order to choose the best model for the task of finding, two models were compared - YOLO and R-CNN according to the following parameters:

- Architecture: YOLO is a single-stage architecture, i. e. it detects objects in the entire image at once, while R-CNN is a multi-stage architecture, i. e. it first extracts regions containing objects and then detects them.
- Speed: YOLO is generally faster than R-CNN because it performs computations on the entire image at the same time without requiring additional computations on regions that do not contain objects.
- Accuracy: R-CNN usually has higher object detection accuracy because it can use more sophisticated methods to identify regions of images that contain objects. However, due to the more complex architecture, it works slower than YOLO.
- Computational resource requirements: YOLO generally requires less resources than R-CNN because it has fewer layers and operations. This makes it more popular for use on resource-constrained devices such as mobile phones.
- Video processing: YOLO usually works better with video because it can identify objects in each frame at the same time. R-CNN, on the other hand, requires additional time to process each frame.

In addition, there are many ready-made solutions for training YOLO models that simplify this task, so it was chosen as the main model for finding objects. At the time of development, there were two most popular YOLO models - v4 and v5. Both versions of the model have advantages and disadvantages:

- Architecture: YOLOv4 has a more complex architecture compared to YOLOv5, which has a simpler and more efficient structure.

- Speed: YOLOv5 is generally faster than YOLOv4 because it has fewer computations and layers.
- Accuracy: YOLOv4 generally has higher object detection accuracy because it uses more sophisticated training methods such as Scaled-YOLOv4 and YOLOv4-P5. However, due to the more complex architecture, it works slower than YOLOv5.
- Image processing: YOLOv5 has better object detection accuracy in small images because it uses high-resolution image processing techniques. YOLOv4, on the other hand, tends to retrain on small images.
- Computational resource requirements: YOLOv5 generally requires less resources than YOLOv4 because it has fewer layers and operations. This makes it more popular for use on resource-constrained devices such as mobile phones.

YOLOv5 training requires a special data structure. An example of such a structure is shown in Figure 5.

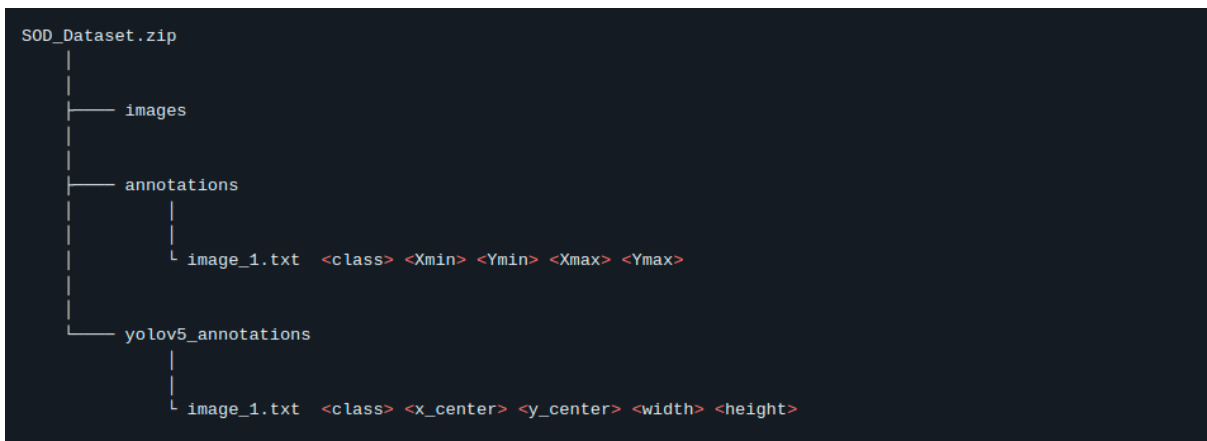


Figure 5: An example of a data structure for model training

The comparative result of model training is shown in Figure 6.

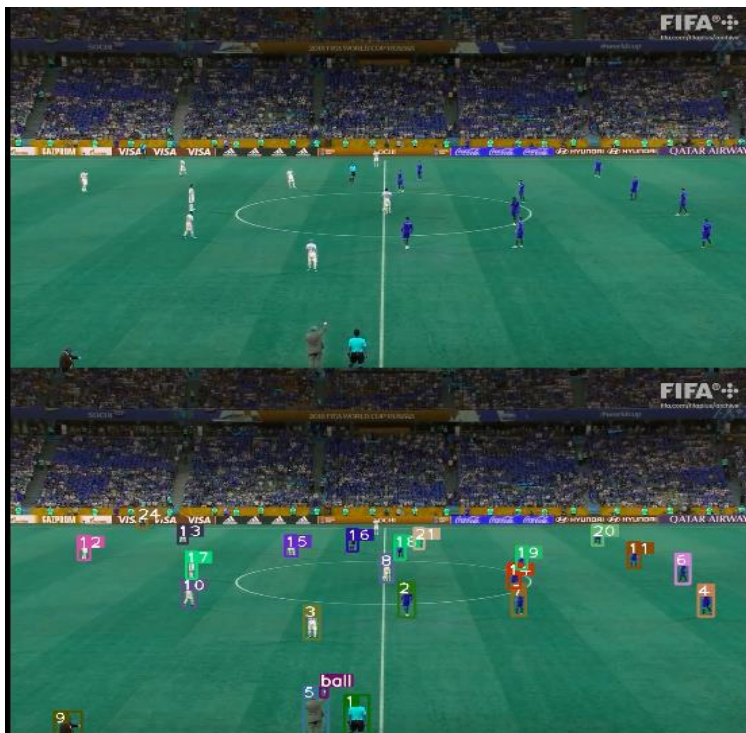


Figure 6: The result of the model of finding players and the ball

The image above shows what the video looks like before processing with the model, and the image below shows what it looks like after.

4.3 Determining the position of the players and the ball on the pitch

To determine the position of the players on the pitch, you first need to find the homography matrix. An algorithm was chosen that uses synthetic data for training for this aim, because its accuracy is sufficient for the given task. An example of the algorithm is shown in Figure 7.

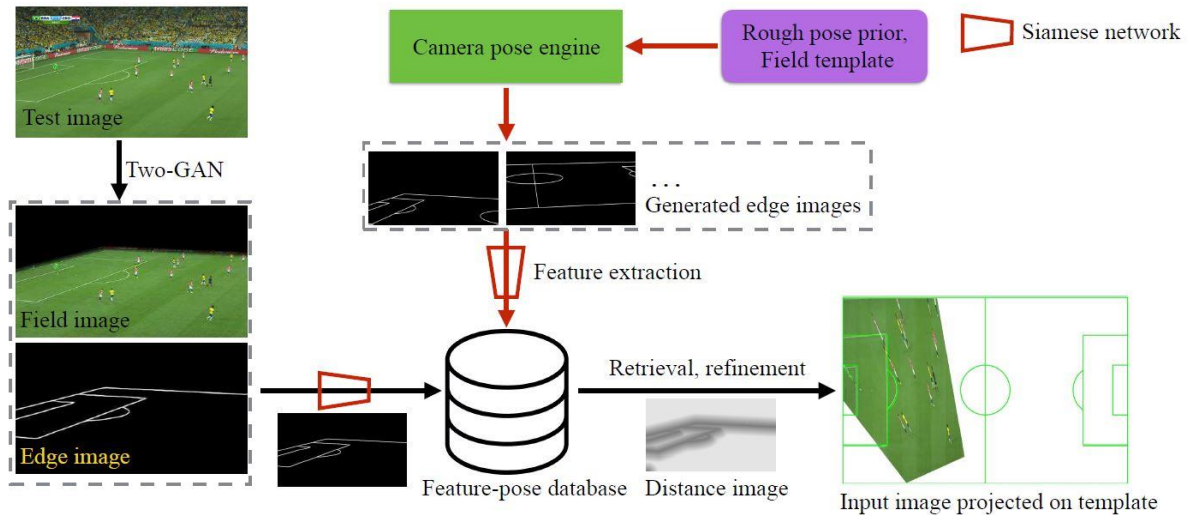


Figure 7: An example of the operation of the homography determination algorithm

The Two-GAN model consists of two pix2pix models – one of which is responsible for the segmentation of grass on the pitch, and the other one is responsible for finding lines on the pitch. Pix2pix is a deep learning model used to generate images using conditional GANs (generative adversarial network). It is capable of generating high-quality images that match the input data.

The pix2pix model uses a pair of images – input and output – for training. For example, the input can be a black and white image, and the output can be a color image. The model is trained to find the relationship between input and output images using conditional GANs. An example of the operation of such a model is shown in Figure 8.

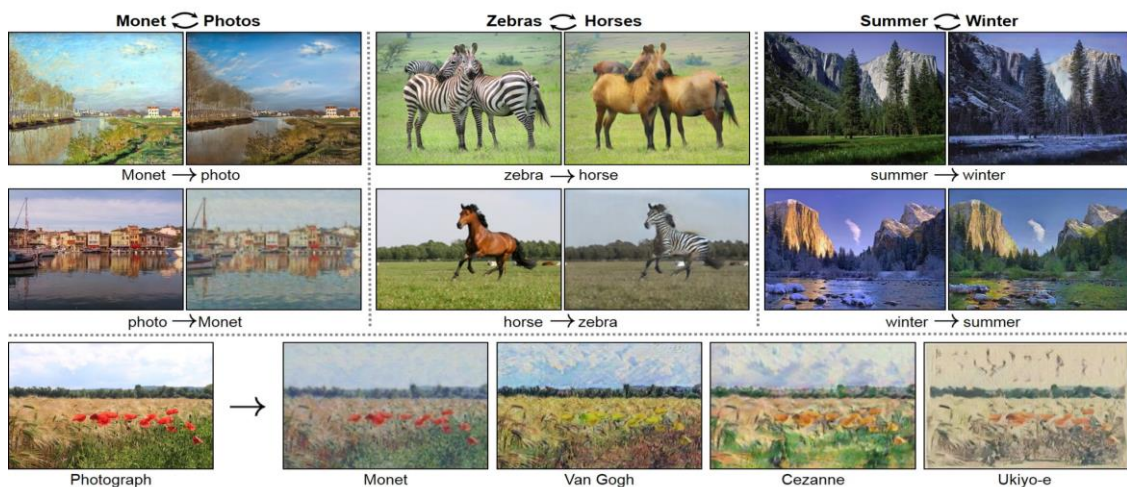


Figure 8: An example of the pix2pix model

In the training process, pix2pix generates images from stochastic noise, which are then transmitted to a discriminator that recognizes whether the image is plausible. Separately, the image is transmitted to the input of the generator, which creates a new output image based on the input image and the learned dependencies. The pix2pix model is very flexible as it can be used for many image generation tasks, such as creating photorealistic images from artificial descriptions, transforming image styles, generating a city map from a satellite image, and more.

However, for the pix2pix model to work successfully, a large amount of training data and high-power computing resources are needed, especially when using large images. An example of a trained model combined with a player and ball location model is shown in Figure 9.



Figure 9: The result of combining the player finding model and the homography matrix definition model

As you can see from the images, the algorithm performs quite accurately, considering that only one camera is used.

4.4 Re-identification of players

To re-identify players, Sportsreid technology is used, which is specifically trained to re-identify players on different frames. Sportsreid contains several different types of models. Comparative characteristics of the models are shown in Table 1.

Table 1
Comparative characteristics of models

Name	Size	Resolution	mAP	rank-1
ResNet50-fc512	24.6M	256x128	81.8	76.1

Name	Size	Resolution	mAP	rank-1
OSNet_x1_0	2.2M	256x128	83.4	78.0
DeiT-Tiny/16	5.5M	224x224	82.2	76.2
DeiT-S/16	21.7M	224x224	84.3	79.4
ViT-B/16	57.7M	224x224	86.0	81.5
ViT-L/16*	303.6M	224x224	89.8	86.7

For comparison, such metrics as mAP and rank-1 are used. mAP (mean average precision) and rank-1 are metrics used to evaluate the effectiveness of computer vision algorithms in the tasks of recognizing objects or people in images or videos. mAP is a metric that measures the average accuracy of object recognition in an image. It takes into account both the accuracy of the found objects and their number. Usually, to calculate mAP, first the threshold for marking an object as found or not found is determined, then the recognition accuracy (precision) is calculated and it is checked whether the correct number of objects were found. The final result is calculated as the average accuracy value for each threshold.

Rank-1 is a metric that measures the accuracy of recognizing people in an image or video. It indicates what percentage of people were recognized correctly when comparing them with the database. Ranking takes place using different algorithms, for example, using the Euclidean distance method or the cosine similarity method between vectors of facial features.

Typically, these metrics are used to compare the performance of different algorithms and models for object or person recognition, helping researchers figure out which algorithm is most effective for a particular task. According to the table, the most optimal option is a model with 2.2m parameters and an image size of 256x128.

Two frames with a difference of 100 shots were selected to test the re-identification performance. For better accuracy, 10 neighboring frames were selected and object detection with tracking via DeepSort was used. After identified unique personalities from DeepSort, next, all found football players were clipped and converted to vectors using Sportsreid. After that, the distance between all vectors was calculated using the cosine distance formula. An example of how the cosine distance formula works is shown in Figure 10.

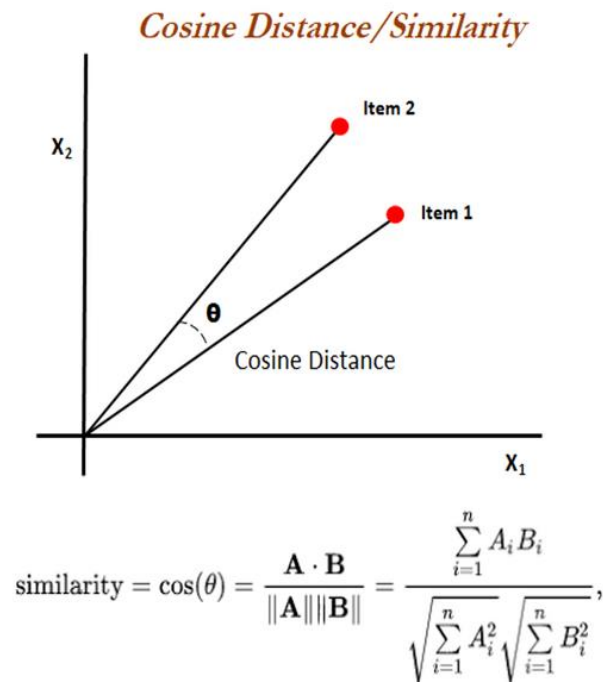


Figure 10: An example of cosine distance operation

After finding the distance, the best candidate among all is selected and it is checked whether the distance is less than a certain threshold. If so, the candidate remains and the process continues. The result of finding candidates is shown in Figure 11.



Figure 11: Candidates and their counterparts

The respondents are shown above, and their candidates below. In total, you can find 9 correspondences between frames. In general, the accuracy of the technique is sufficient for the initial tasks, but it needs to be refined in the future.

4.5 Calculation of statistical data

For a football player, the importance of different metrics may depend on the role he plays on the pitch.

Time of ball possession. This metric indicates how long a player keeps the ball at his feet. This is important for players who are responsible for deploying the team's attack. The longer a player holds the ball, the more time he has to make the right decision and pass the ball to his partners. To calculate this metric, information about the position of the ball in the frame and the player is used. If the ball is close to the player (30 pixels) or the last coordinates of the ball are the coordinates of the player's feet, then the time of possession of the ball is credited to the player. The frame rate in the video is also taken into account for a more accurate calculation of the metric.

Distance traveled. This metric indicates how many meters the player covered during the match. This is important for players who are responsible for covering a large area on the pitch. Players such as forwards, defenders and midfielders must be responsible for moving from their side to the opposing side and back to help their team in attack and defence. Data to calculate this metric is collected only when the player is visible in the frame. During the video, the homography matrix is calculated, compared with the player's movement, and calculates the approximate number of meters he has traveled.

Number of passes. This metric indicates how many times a player has passed to his partners. This is important for players who are responsible for organizing the team's attacks. Midfield and attacking players usually need to be good passers, as their passes can lead to goals or other scoring opportunities. To calculate this metric, the position of the players and the ball is used. If the ball was close to the player (30 pixels), or the trajectory of the ball started from the feet of one player of the team and passed into the near zone or the feet of another player from the same team, then such an event is counted as a pass.

Number of interceptions. This metric indicates how many times a player stopped an opponent's attack by intercepting the ball. This is important for players who are responsible for the team's defense. Defensive and midfield players usually need to be good interceptors, as their ability to stop opposing attacks can be critical to a team's success. If the ball was close to the player (30 pixels), or the trajectory

of the ball started from the feet of one player of the team and passed into the near zone or the feet of another player from the opposite team, then such an event is counted as an interception.

Time Spent on Own Side and Time Spent on Opponent's Side: These metrics indicate how much time a player has spent on their own side and on the opponent's side. This is important for all players as it helps to understand where and how a player spends his time on the pitch and how it can be used to help the team succeed. To calculate these metrics, information about the position on the pitch is used, which is obtained using the homography matrix.

Overall, these metrics help players and coaches analyze and improve the performance of players and the team as a whole. And if a player can improve in one of these metrics, it can have an immediate impact on the team's success.

4.6 Application Programming Interface (API)

API (Application Programming Interface) is a set of protocols, tools and standards used to develop software and ensure interaction between various software components.

An API defines how different software components should interact with each other and what actions and operations can be performed from these components. An example of returning API information is shown in Figure 12.

```
○ ○ ○
{
  "players": [
    {
      "uid": "player1",
      "team_id": "team1",
      "owning_ball_time_s": 240,
      "walked_distance_km": 7.2,
      "passes": 45,
      "interceptions": 3,
      "own_side_time_m": 30,
      "opposite_side_time_m": 20
    },
    {
      "uid": "player2",
      "team_id": "team1",
      "owning_ball_time_s": 180,
      "walked_distance_km": 6.5,
      "passes": 32,
      "interceptions": 5,
      "own_side_time_m": 25,
      "opposite_side_time_m": 18
    },
    {
      "uid": "player3",
      "team_id": "team2",
      "owning_ball_time_s": 210,
      "walked_distance_km": 8.1,
      "passes": 39,
      "interceptions": 2,
      "own_side_time_m": 28,
      "opposite_side_time_m": 22
    },
    ...
  ]
}
```

Figure 12: Example of information returned from the API

The application API is implemented in Python using Flask. Flask is a lightweight framework for building web applications in Python. API in this case implements the following functions:

- **Receive video:** The API can receive video that needs to be decoded for further processing.
- **Video Processing:** Encoded video is processed using machine learning models to derive metrics and player data. These metrics include time of possession, distance covered, number of passes, number of interceptions and other characteristics that are important.
- **Returning results:** The last step of the API is to return the results of video processing as a JSON object.

The overall API architecture includes components such as router, machine learning model, database, and others.

5. Conclusions

In this article, a methodology for the formation and processing of statistical information about sports matches based on the use of neural networks was developed. The Yolov5 model was used to locate the players and the ball in the image. It is fast and accurate, so it is ideal for such a task. To determine the position of the players on the pitch, pix2pix neural networks were used. They were trained on images of players and their positions on the pitch, allowing them to accurately identify players' positions in new images.

Soccerreid technology was used to re-identify players. It allows you to distinguish one player from another based on their appearance and movements. Also, the creation of an API in Python using Flask to obtain statistical information about players and their actions on the pitch was described. This allows coaches and analysts to obtain valuable information to improve team strategy and tactics.

Further research may concern improving the accuracy of the developed methodology, its speed and enrichment of statistics. The scientific novelty of the obtained results lies in the fact that the developed automated method of generating statistical data has a great potential for application in the industry of football match analysis. The results can be used for further research in the field of automating the formation of statistics. The obtained results can also be used to improve work in other areas related to sports.

6. References

- [1] J. Assfalg, M. Bertini, C. Colombo, A. Del Bimbo, W. Nunziati, Semantic annotation of soccer videos: automatic highlights identification, *Computer Vision and Image Understanding* (2003) 285–305. doi: 10.1016/j.cviu.2003.06.004.
- [2] Ch. Perin, R. Vuillemot, J.-D. Fekete, SoccerStories: A Kick-off for Visual Soccer Analysis, *Proceedings of the IEEE Transactions on Visualization and Computer Graphics* 19(12):2506-15, Dec. 2013. doi:10.1109/TVCG.2013.192.
- [3] M. Stein, H. Janetzko, A. Lamprecht, T. Breitzkreutz, P. Zimmermann, Goldlucke, B., Schreck, T., Andrienko, G., Grossniklaus, M. and Keim, D. Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics* (2018) 13–22. doi: 10.1109/TVCG.2017.2745181.
- [4] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* (2004). URL: [ijcv04.pdf\(ubc.ca\)](http://ijcv04.pdf(ubc.ca)).
- [5] S. Afzal, S. Ghani, M. M. Hittawe, S. F. Rashid. Visualization and Visual Analytics Approaches for Image and Video Datasets: A Survey, *The ACM Transactions on Interactive Intelligent Systems* 13 (2023). doi: 10.1145/3576935.
- [6] N. Ding, W. Jin, K. Takeda, Y. Bei, K. Fujii. Estimation of control area in badminton doubles with pose information from top and back view drone videos, *Multimedia Tools and Applications* (2023). doi: 10.1007/s11042-023-16362-1.
- [7] Z. Chen, Q. Yang, J. Shan, T. Lin, J. Beyer, H. Xia, H. Pfister. iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations, *Human-Computer Interaction* (2023). doi: 10.48550/arXiv.2303.03476.
- [8] M. B. Jurca. A modern approach for positional football analysis using computer vision, *Proceedings of the 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sept. 2022. doi: 10.1109/ICCP56966.2022.10053962.
- [9] S. Rahimi, A. Moore, P.A. Whigham. A vector-agent approach to (spatiotemporal) movement modelling and reasoning, *Scientific Reports* (2022). doi: 10.1038/s41598-022-22056-9.
- [10] Z. Chen, J. Beyer, H. Pfister, Q. Yang, H. Xia, X. Xie, Y. Wu. Sporthesia: Augmenting Sports Videos Using Natural Language. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Oct. 2022. doi: 10.1109/TVCG.2022.3209497.

- [11] J. Wang, K. Hu, Zh. Zhou, J. Ma. Tac-Trainer: A Visual Analytics System for IoT-based Racket Sports Training, *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Oct. 2022. doi: 10.1109/TVCG.2022.3209352.
- [12] X. Xie, Y. Wu, D. Deng, Y. Wu. OBTracker: Visual Analytics of Off-ball Movements in Basketball. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Sept. 2022. doi: 10.1109/TVCG.2022.3209373.
- [13] B. Jackson, T. Y. Lau, D. Schroeder, KC. Jr. Toussaint, D. F. Keefe. A lightweight tangible 3D interface for interactive visualization of thin fiber structures. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Dec. 2013. doi: 10.1109/TVCG.2013.121.
- [14] K. Moreland. A survey of visualization pipelines. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Mar. 2013. doi: 10.1109/TVCG.2012.133.
- [15] S. Rahimi, A. B. Moore, P. A. Whigham, A vector-agent approach to (spatiotemporal) movement modelling and reasoning, *Scientific Reports* (2022). doi: 10.1038/s41598-022-22056-9.
- [16] J. Beernaerts, B. De Baets, M. Lenoir, N. Van de Weghe. Qualitative Team Formation Analysis in Football: A Case Study of the 2018 FIFA World Cup, *Frontiers in Psychology* (2022). doi: 10.3389/fpsyg.2022.863216
- [17] A. Benito Santos, R. Theron, A. Losada, J. E. Sampaio, C. Lago-Peñas. Data-Driven Visual Performance Analysis in Soccer: An Exploratory Prototype. *Frontiers in Psychology* (2018). doi: 10.3389/fpsyg.2018.02416.
- [18] C. D. Stolper, M. Kahng, Z. Lin, F. Foerster, A. Goel, J. Stasko, D. H. Chau. GLO-STIX: Graph-Level Operations for Specifying Techniques and Interactive eXploration. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Dec. 2014. doi: 10.1109/TVCG.2014.2346444.
- [19] F. Lord, D. B. Pyne, M. Welvaert, J. K. Mara. Capture, analyse, visualise: An exemplar of performance analysis in practice in field hockey, *PLoS One* (2022). doi: 10.1371/journal.pone.0268171.
- [20] P. Isenberg, P. Dragicevic, W. Willett, A. Bezerianos, J.D. Hybrid-image visualization for large viewing environments. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Dec. 2013. doi: 10.1109/TVCG.2013.163.
- [21] X. Hu, L. Bradel, D. Maiti, L. House, C. North, S. Leman. Semantics of directly manipulating spatializations. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Dec. 2013. doi: 10.1109/TVCG.2013.188.
- [22] M. Krzywinski, I. Birol, S. J. M. Jones, and M. A. Marra. Hive plots: a rational approach to visualizing networks. *Briefings in Bioinformatics*, 13(5):627–644, Sept. 2012. doi: 10.1093/bib/bbr069.
- [23] I Konovalenko, P Maruschak, V Brevus. Steel surface defect detection using an ensemble of deep residual neural networks, *Journal of Computing and Information Science in Engineering* 2022, 22(1), 014501, <https://doi.org/10.1115/1.4051435>.
- [24] I. V. Lytvynenko; P. O. Maruschak; S. A. Lupenko; Yu. I. Hats; A. Menou; S. V. Panin. Software for segmentation, statistical analysis and modeling of surface ordered structures, *AIP Conf. Proc.* 1785, 030012, 2016, <https://doi.org/10.1063/1.4967033>.
- [25] 3. Shymchuk, G., Lytvynenko, I., Hromyak, R., Lytvynenko, S., Hotovych, V. Gas Consumption Forecasting Using Machine Learning Methods and Taking into Account Climatic Indicators. *CEUR Workshop Proceedings*, 2023, 3468, pp. 156–163.