



## AI FOR BETTER STORYTELLING IN LIVE FOOTBALL

N. Déal<sup>1</sup> and J. Vounckx<sup>2</sup>

<sup>1</sup> UEFA, Switzerland and <sup>2</sup> EVS, Belgium

### ABSTRACT

Artificial Intelligence (AI) represents almost limitless possibilities for the future and is already having a transformational impact in many areas. Live football production is one where there is a real need to drive operational efficiencies as rights holders are pushed to deliver more and better live content to an increasingly diverse and connected audience.

With the right machine learning agents in place, AI is able to mimic creative human behaviour, overcoming the limitations of automation and performing high-level, complex tasks with the speed and reliability that is required in live environments.

In this paper, we will present our vision of AI as a built-in production assistant, through some potential future applications as well as current use cases in the context of live football storytelling. We will showcase applications such as assisted framing, camera selection, camera calibration and automatic robotic camera steering that all contribute to delivering better storytelling in live sport productions.

### INTRODUCTION

Viewers' expectations are on the rise. Modern technologies are changing the way football fans are watching their favourite sport, increasing the need for more immersive experiences that make them feel closer than ever to the game and the players.

Already, leagues, teams and individual sporting events are capitalising on this new reality, exploring the latest technologies that help them to produce compelling content that suits the desires of today's digital-savvy fans.

However, there is one major catch – this needs to happen with increasingly tight budgets. Broadcasters are expected to bring more efficiencies to their productions, while ensuring they are offering a superior viewing experience for their audiences.

In this context, AI-based technologies are progressively finding their way into the broadcast industry for live productions. Based on various machine learning components, AI's ability to mimic the behaviour of human operators and their creativity, is opening the doors to a whole new world of opportunities.

Johan Vounckx, SVP Technology & Innovation from EVS, and Nicolas Déal, TV Transmission Manager at UEFA, explore different use cases of AI in football storytelling, that may drive the development of live production technology in the years ahead.



## **AI ON THE FOOTBALL PITCH**

### **The Attractive Power of AI**

With the recent advances in neural network technology, and more specifically the ability to execute deep neural networks in real-time, AI has opened the door to a new kind of creative automation that lends itself well to live productions (1). It allows for the replication of the human artistic touch and the ability to cope with unforeseen events (2).

Neural network technology solves problems not by the explicit, rigid programming of a solution, but by learning from a large set of examples (3). By way of an example, say a set of images needs to be sorted into those with and those without cars in them (4). This image recognition problem can be solved by a neural network which will be given the images, along with the correct answers of 'car' or 'no car'. It uses these to learn the difference between the two kinds of images. Based on those inputs, the network can learn how to respond to new images, telling us whether or not they contain a car.

This process is similar to us humans, who learn by example and adjust our behaviour based on new experiences. After several attempts, we improve and eventually become competent, until we become an expert at addressing a specific challenge or 'input' (2).

Because it is still early days, it is difficult to foresee how AI will impact the future of live productions. But one certainty is that human oversight will always be important (5). Indeed, the creativity and flexibility of human operators will remain central for the success of live productions, and the machine learning process will always require the input of a human.

That said, we can easily envision a future where AI acts as an enabler - or an assistant - for the completion of certain tasks in live productions and deliver smarter workflows that would allow production teams to free up their time for more creative tasks.

### **AI as a virtual production assistant**

Again, compelling live storytelling cannot happen without the human touch. The best stories arise from a human's creativity and their emotional intelligence. However, unlike humans, machines do not get bored, and they do not need to take a break to remain efficient. In certain circumstances, humans can be slower and less regular in the performing of their tasks, and this is where it makes sense for AI to intervene.

With the right machine-learning neural networks in place, AI can perform multiple tasks with the speed, reliability and consistency that are required within live productions – and that humans sometimes lack.

We present the concept of AI used as a built-in virtual production assistant, where operators and directors are assisted in the performing of their tasks by a system that is powered by a series of real-time engines (analysis engine, A/V processing engine, and a content creation engine). The analysis engine analyses in real-time the multiple audio, video and other data feeds that are generated during the production, or that can come from other sources such as social media or archives. The result of this analysis is a set of metadata such as events that happened (a red card, a goal kick, ...), the indication of objects in the video, or the indication of the heat of the action in the image. The most immediate use of such metadata is the automatic generation of logging information during the matches.

This metadata is both stored for later processing and used immediately by the other real-time engines. The A/V processing engine uses the metadata to create audio and video material that can be used in the production. A simple example is the graphical insertion of information in the video (such as the indication of a fault, or the indication of biometric data of the players, or the drawing of

an offside line). The A/V content that is generated is then used in the regular production workflow. A last engine is the content generation engine. This is an engine that analyses metadata in realtime to auto-generate content. As an example, one can think of the automatic generation of highlights, or the automatic cropping of images. In both cases, the content generation engine creates the instructions to generate the A/V content. The actual processing of these instructions to actually create an A/V feed, is done by the A/V processing engine.

The system functions in two modes. The first being an automatic mode where suggested content (proposed replay sequences or proposed camera angles for instance) is pushed directly to the operators. Clearly, in this mode, the content generation engine is heavily used. This allows the directors to cope with the increasing complexity of today's productions and to react faster since the AI system pre-filters and pre-suggests what it deems the best content to be shown. An extreme case is a fully automatic production. Whereas this is not envisaged for major events, a fully automatic production is very relevant for long tail content, such as youth games or local matches, where traditional production methods are not realistic for budgetary reasons.

The second is an operator-steered mode where an AI-based natural speech processor interprets requests given by operators and directors such as 'give me a camera angle that shows the goal', 'give me a slow-motion view of this clip', or 'give me a replay of the last goal'. These requests are then translated by the natural speech processor into API instructions destined for each AI engine, that in turn, produce the desired A/V outputs.

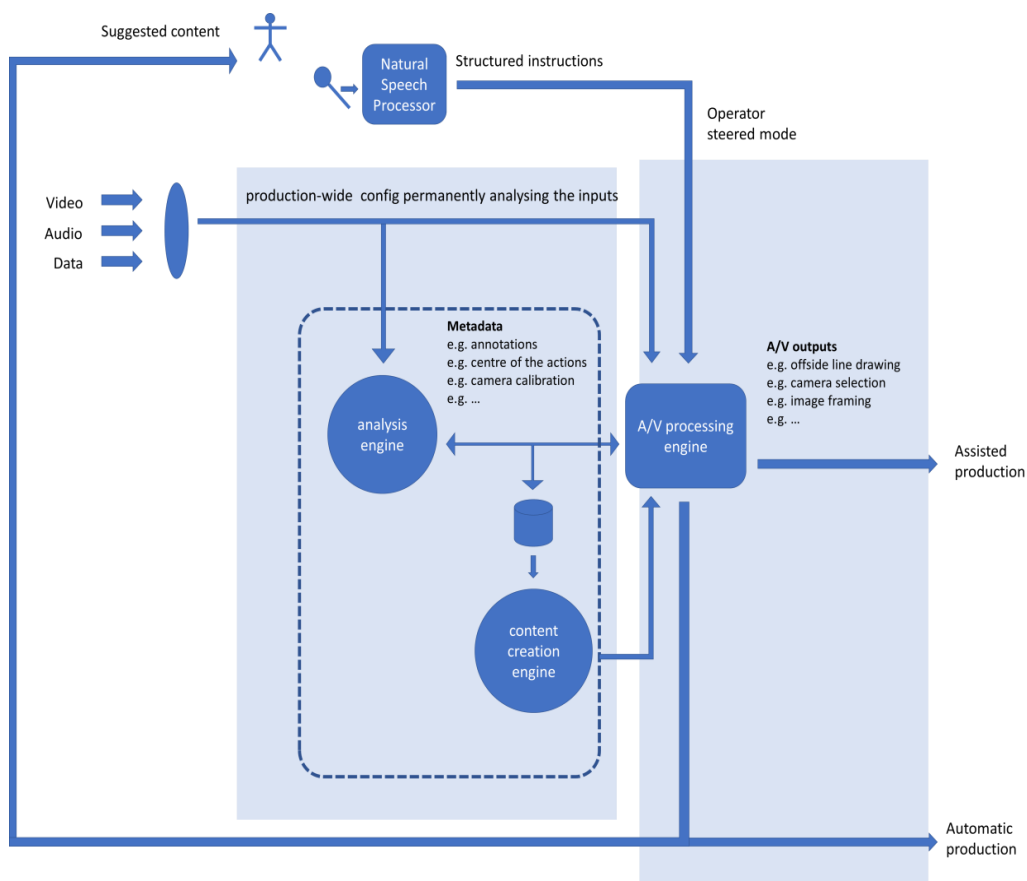


Figure 1 - the virtual production assistant system.



We've identified a wide number of application areas where the concept of AI as a virtual production assistant system can be envisaged:

### **Anticipated application areas**

#### AI for live data analysis and the automation of certain tasks:

- Logging and indexing
- Camera calibrations
- Player and object tracking
- Performing intelligent searches for archival assets to integrate into the current broadcast (6)
- Interpreting and predicting game behaviour

#### AI for the preparation of video sequences to be aired upon request or as a proposal:

Here, two use cases can be envisaged, based on spatial and temporal selection of the existing A/V sources:

- Generating camera viewpoints & camera angles (either as a smooth evolution of the camera viewpoint, either by switching between viewpoints)
  - o automatic camera viewpoints
  - o viewpoints upon request (e.g. "give me a viewpoint with Messi")
- Generating replays & highlights
  - o automatic camera viewpoints
  - o viewpoints upon request (e.g. "give me the latest fault")

#### AI for the preparation of high-quality output:

- Graphic overlays in the correct visual perspective, leveraging the camera calibration performed by the analysis engines
- Creating slow motion clips from regular cameras
- Performing the colour shading
- Presenting additional camera views obtained via a recombination of existing and interpolated camera images
- Automatic steering of robotic cameras

In addition to the abovementioned application areas, we see other specific use cases where live football productions would benefit from the integration of an AI production assistant:

#### AI to cut and propose replay clips:

Instant replays are so commonly used in the live production of sports events, they've become a fundamental part of the storytelling and it's hard to imagine watching a game without them. However, one of the issues with replays is that they consume airtime at the expense of the live action coverage. To remediate this situation, an AI-based system could cut the replay clips then propose them directly to the viewers on different screens, as soon as they are ready. It would be up to the viewers to decide how many replay clips they would like to watch, and most importantly, when. The clips would show up on the main window, but the live stream would still be visible, offering the possibility to come straight back to the field, would some eye-grabbing action occur.

Figure 2 below shows how an AI-based solution could present replay clips directly to the viewer's iPad. As soon as the various replays are produced, they would show up below the live match, allowing the user to switch between the different replays and the live match at their convenience.

The AI assistant would automatically detect the aspect ratio of the viewer's device and crop the format accordingly.



Figure 2 – replay clips are proposed automatically to the viewer, with the live stream still visible.

### AI for live coverage of the action using fixed cameras:

With AI, the numerous HD mobile cameras that are found in the stadium could be replaced by fixed 8K cameras that are strategically placed to cover the entire football pitch, reducing the amount of equipment needed. With the right deep learning algorithms, the AI assistant would be able to detect and extract the relevant action. The resulting data stream of each camera would then be transmitted to a data centre for the pictures to be processed, stored or used as a live stream in a live production.

The vision we have for the future of live productions is one that combines the best of both worlds: the speed and reactivity of AI to increase operational efficiencies by taking on tasks that involve labour-intensive content/data management (7) - and the flexibility and creativity of humans, who remain in the driving seat and are able to focus on their primary activity: to create the most compelling stories for a top-notch viewing experience.

In the next section, we will provide some examples of how EVS is working to integrate AI into live football productions.

## **THE WARM-UP TO AI-ASSISTED STORYTELLING**

### **Assisted framing**

Assisted framing is a function which uses AI to crop regular HD feeds into smaller aspect ratios to fit smartphone and other screen formats. This is used for the publication of content on social media or on a mobile-first feed, where varying aspect ratios are required.

Conventional approaches such as centre-aligned cropping or object detection have their limitations: the action is rarely in the middle of the original camera image, so you face the issue of missing a big part of the action by choosing centre-aligned cropping. And although object detection might seem like a better option, it is not easy to put into practice since the ball can get hidden behind players or have other objects mistaken for it. Furthermore, the ball is not always the most interesting part of the action, as other actions can distract from the ball, such as the players.

With AI, the system can extract the key element in the image and mimic the behaviour learnt from the training set, where learning data is shown to the network after a human operator indicates the best optimal focal point of the action from a variety of games.

Technically, the assisted framing is based on a sequence of real-time engines. The first engine, based on a neural network capable of identifying the focal point of the action in the different images

of the video, returns the Region of Interest for each individual image in the video sequence. This Region of Interest is identified by a set of values, including its position  $(x_{rawi}, y_{rawi})$  corresponding to the current image (with index  $i$ ) in the video. These values are forwarded to a second real-time engine that will apply a temporal filtering. The temporal filtering ensures a smooth evolution of the Region of Interest, avoiding too sudden jumps in the resulting framed video. In order to apply this filtering, the real-time engine uses the Regions of Interest of the previous images in the video, including the positions  $\{(x_{rawi-1}, y_{rawi-1}), \dots\}$ . This results in a tuned Region of Interest for the current image. Finally, a real-time video processing engine will cut out the desired Region of Interest from the original image and will forward that image to subsequent stages in the production workflow.

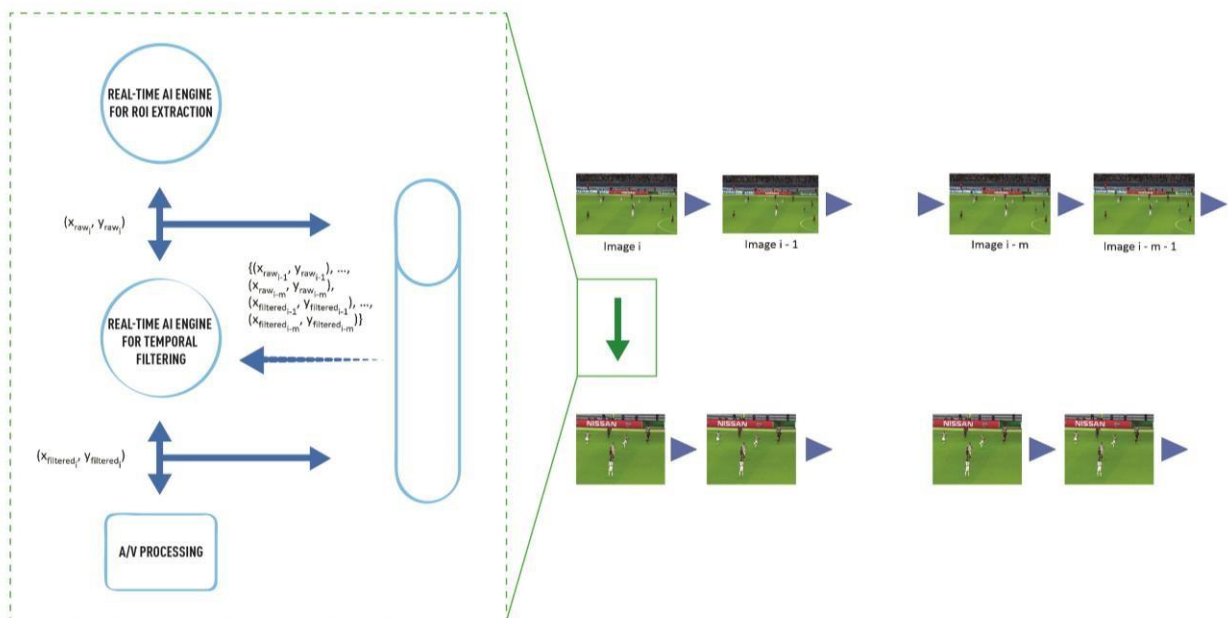


Figure 3 – assisted framing approach based on a sequence of real-time engines

Using this method, the sequence in figure 4 below, shows how AI can create an ideal framing that goes beyond conventional approaches:

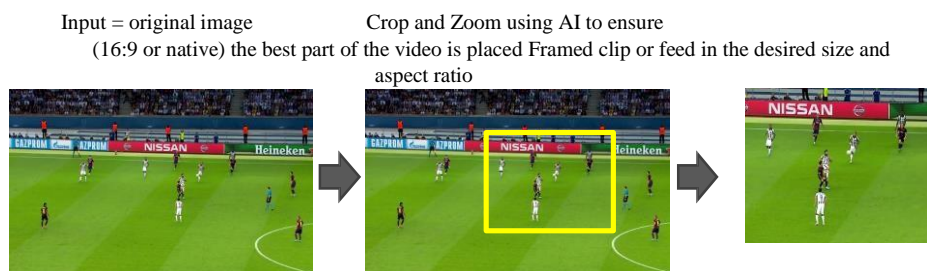


Figure 4 – assisted framing starts from the original image, and, based on deep learning, cuts out a cropped and zoomed part that corresponds to the heart of the action.

### Assisted camera selection

For any important football game, there are multiple cameras in place to make sure the action is covered at all times and it's up to the director to decide which of those many camera angles will be shown on TV. Assisted camera selection uses AI to self-select what it deems to be the best, or most appropriate camera angle. To measure the results, we asked people to rate different clips from the same game. One clip was the human directed clip, the second was the algorithmic solution and a third clip that was semi-randomly generated was added. The results indicate that the machine-directed clips, with the exception of a purposely complex scene, performs at the same level as the human director.

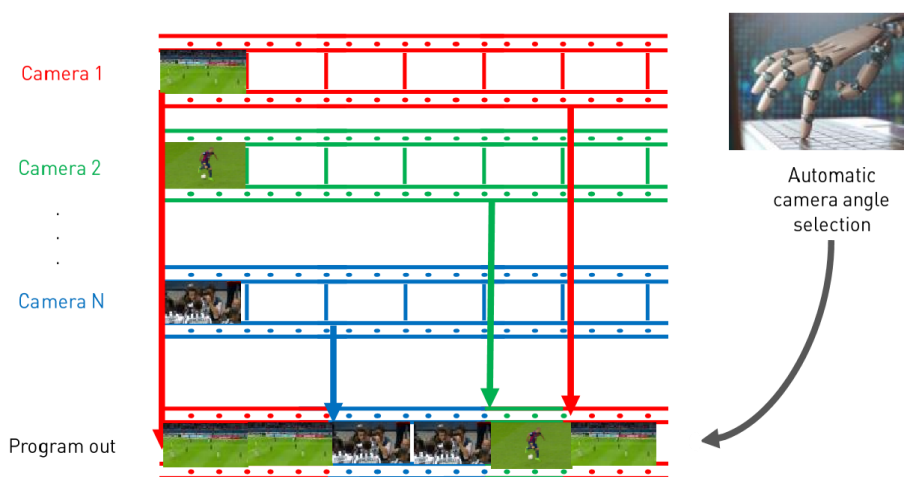


Figure 5 – assisted camera selection – the system automatically selects the best camera angle to create a compelling program view

### Assisted camera calibration

AI can perform camera calibration in real-time based on the video image. For each video image, the transformation between the camera image and the layout of the field is calculated. This allows to easily draw lines and other objects in the rectangular 2D view of the field, and then to project these lines and objects into the real image with the right visual perspective. A neural networkbased approach calculates what a 2D version of the field would look like from an in-stadium camera. Then, realizing the link between the two, elements can be added automatically. The first application of this is the accurate addition of an offside line for use in a football game. With machine learning techniques, the system shows operators the exact position on the field where players would be offside – all shown on the in-stadium camera's video output.

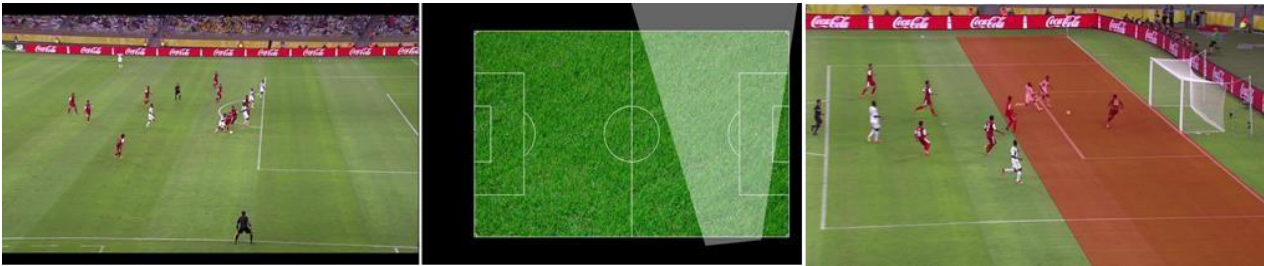


Figure 6 – the automatic calibration creates a link between the camera image and a 2D layout of the pitch. The last image shows how an offside line is then drawn with the right perspective.

The automatic camera calibration is based on an analysis of the images generated by the camera. A concatenation of several AI engines identifies reference markers in the football field and combines these to map the real camera image on a mathematical representation of a soccer field. This mapping is used to extract the parameters (8) that identify a camera calibration and that specify the transformation between the real camera images including the distortion and the optical perspective on one hand, and the real world on the other hand. Based on these calibration parameters, overlaying objects on the real video image boils down to applying the inverse transformation of the real world object in a real world coordinate system into the coordinates of the obtained video images.

### **Assisted robotic camera steering**

Technologies comparable to AI-assisted framing open the door for other applications including automatic robotic camera steering. With this approach, AI predicts the action and moves the camera into the right direction to ensure that it follows what it thinks is the most relevant part of the image. Figure 7 shows how the system analyses the scene from a wide-angle camera covering the complete pitch to detect the relevant information. The AI module then uses that information to steer each robotic camera by P/T/Z commands, pointing the camera in the right direction. This method generates interesting and dynamic viewpoints - in real-time and with the appropriate zoom factor.

The automatic robotic camera steering builds on a number of AI engines. A first component, applied at the beginning of the production or whenever the reference position of the cameras changes, uses an AI engine to automatically calibrate the cameras, allowing to correlate the different camera locations and positions inside the images created by these cameras. The calibration is based on an analysis of the images created by the cameras.

A second component will ensure a real-time steering of the different robotic cameras. The images coming from a camera covering the complete soccer field, are fed into an artificial intelligence engine that will define the area(s) that should be filmed by the robotic cameras. These area(s), thanks to the calibration, are translated into concrete P/T/Z instructions to the robotic cameras. The images created by the robotic cameras are processed by regular production equipment, independently from the AI engines.

The glass-to-glass latency between the camera capturing the complete soccer field, and the robotic camera pointing to the appropriate area of the field as indicated by the AI engine, is very critical. If this latency becomes too high, the robotic cameras will never follow the real action and will always lag behind the ideal position. A lot of effort has been spent in optimising this end-to-end latency, both by optimising the raw performance and transfer speeds of the processing pipeline, and by tuning the AI algorithms.



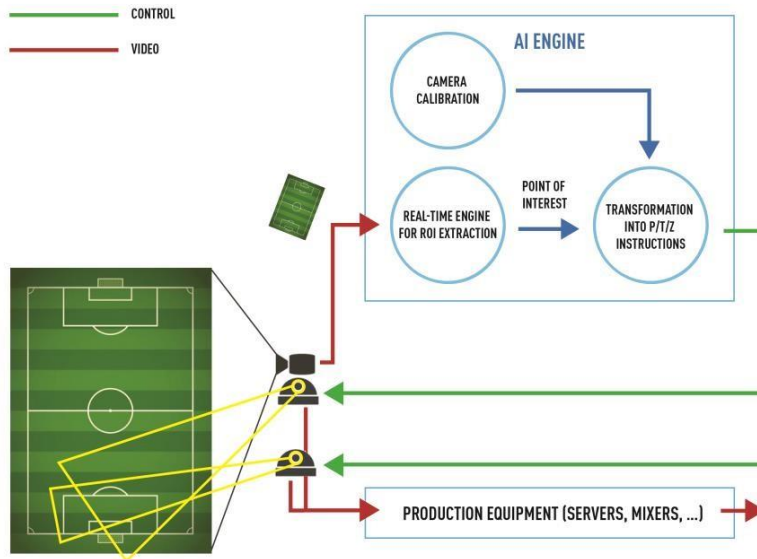


Figure 7 – robotic camera steering system

### Assisted slow motion

Slow motion replays are extremely valuable in football and in sports in general as they allow viewers to appreciate the skills of the players or athletes and help audiences better understand a given action or incident. Viewers particularly enjoy the clean image quality from super motion cameras that make the experience even better for storytelling. However, because super motion cameras are expensive, they are limited to very few, strategic camera positions.

With AI however, it becomes possible to create high-quality super motion images from regular cameras. Neural networks can be trained to create video interpolation, where virtual, intermediate frames are inserted in between the real, physical images for a much higher framerate video.

Figure 8 shows how AI-generated super motion creates higher-quality replays in comparison with super motion created by repeat frames - rather than repeating frames, the creation of additional frames generates smoother, more coherent video sequences.

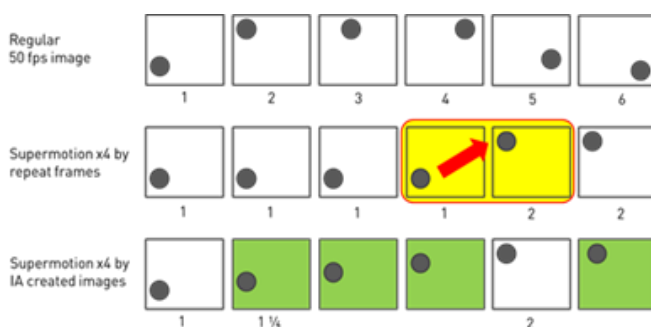


Figure 8 – smoother super motion replays with AI-created images



This method could be extremely beneficial for smaller-scale productions for instance, that do not have the financial power to deploy super motion cameras. It also allows to create ultra motion or even hyper motion replays that can be streamed in real time from the camera. Another benefit is the possibility to create such replays on existing footage, and integrate them into the live broadcast, adding greater production value to live events.

## **CONCLUSION**

Far from aiming to replace people with machines, AI in the broadcast industry is being developed to help humans do their job faster and more efficiently. Based on deep learning methods, AI can support operators and directors by automating certain tasks, analysing video and preparing A/V assets to be shown in real-time, whether that be in the form of replays, highlights, or a compelling selection of camera angles, and even the creation of complementary content. We envision a future where AI is used in live football productions to enable operators and directors to cope with the increasing complexities of live workflows, and help them create more engaging storytelling, meaning more fan loyalty and engagement over time.

## **REFERENCES**

1. Grotticelli M., April 2018, "At NAB 2018 artificial intelligence touted as super-charged video assistant", [thebroadcastbridge.com](http://thebroadcastbridge.com).
2. Magera F., Vounckx J., April 2018, "How AI will take productivity in the broadcast industry to the next level", NAB 2018 technical paper.
3. Hastie T., Tibshirani R., Friedman J., 2009, "Overview of Supervised Learning". In: *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY
4. Krizhevsky A., 2009, "Learning Multiple Layers of Features from Tiny Images"
5. Sacchelli D., February 2018, "How AI will change the broadcasting and entertainment landscape", [itproportal.com](http://itproportal.com)
6. Clevinger D., September 2017, "How AI will disrupt sports entertainment networks", [venturebeat.com](http://venturebeat.com)
7. Alamares M., October 2017. "AI will soon bring huge changes to live video production", [streamingmedia.com](http://streamingmedia.com)
8. Hartley R., Zisserman A., 2003. "Multiple View Geometry in Computer Vision"