EVS Internships 2025-2026

Geometry Vision Transformers Applied to Broadcast Scenarios

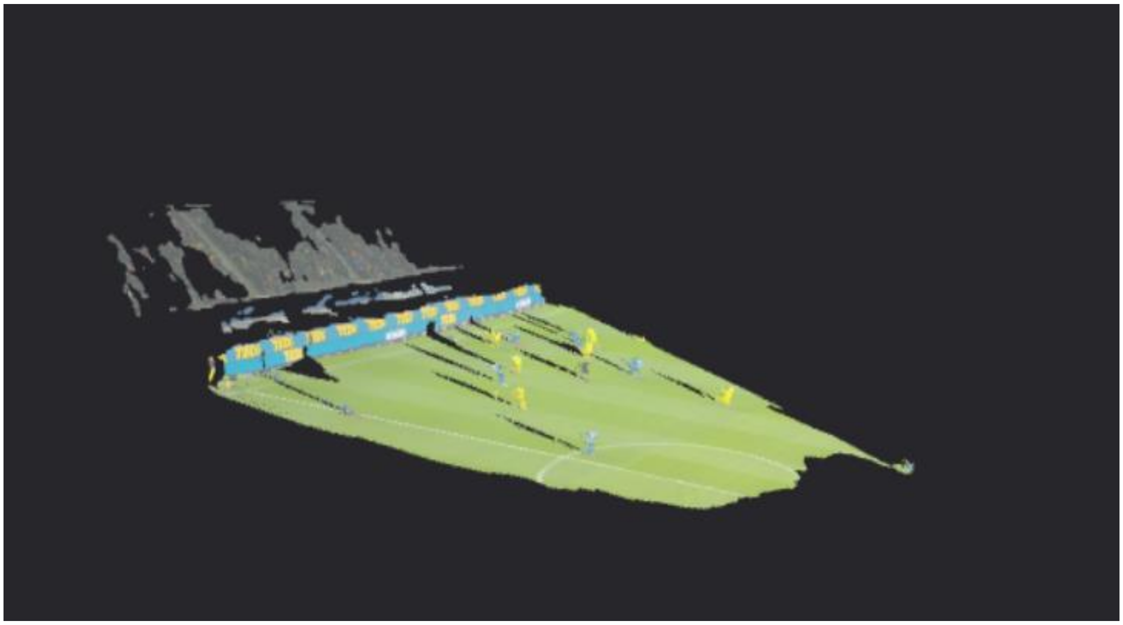
Keywords: computer vision, deep learning, transformers

The past year has seen several breakthroughs in terms of 3D reconstruction from images. Until recently, the best methods for 3D reconstruction were classical SfM methods such as COLMAP[1], which performs feature detection and matching, camera calibration, triangulation as sequential steps in an iterative setup.

The latest developments from DUSt3R[2], and alikes as well as VGGT[3] suggest that big transformer models are able to capture the scene geometry from very few images as demonstrated in the figure below, in a single step inference.

The purpose of this internship is to investigate the relevance of such transformers for our applications. Is the quality of the reconstruction sufficient to help novel view synthesis systems from the few and sparse viewpoints obtained by broadcast cameras? What level of precision can we expect from the camera parameters that can be derived from the obtained 3D points?

The ideal candidate for this internship has a good background in deep learning and computer vision, especially on camera calibration, and some experience with python, PyTorch, and Linux.



Point cloud results of VGGT from a single soccer image.

[1] *"Structure-from-Motion Revisited", J. Schönberger et al., CVPR, 2016*
[2] *"DUSt3R: Geometric 3D Vision Made Easy", S. Wang et al., CVPR 2024*
[3] *"VGGT : Visual Geometry Grounded Transformer", J. Wang et al., CVPR 2025*

Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com *and* o.barnich@evs.com

EVS Internships 2025-2026

Computer Graphics to Photorealistic Style Transfer for Sports Broadcasting

Keywords: Computer Vision, Deep Neural Networks, Diffusion Models, Domain Adaptation, Computer Graphics

Recent deep learning techniques bring us closer and closer to the ultimate live sports 3D replay technology that would allow replay operators to operate a photorealistic virtual camera to render the shot of their dreams for any thrilling sports action, breaking free from any camera constraints.

On one side, photorealistic novel view synthesis techniques such as Neural Radians Fields and Gaussian splatting are steadily improving but are still struggling when only having a few cameras around the scenes at hand.

On the other side, skeleton tracking and reconstruction techniques are now production ready. At the same time, 3D rendering ecosystems such as Blender and Unreal Engine allow real time rendering of those generated metadata to create virtual cameras. However, the quality of such rendering are still not quite photorealistic.

Style transfer techniques however offer a promising alternative to transform geometrically perfect CG renderings into great photorealistic virtual cameras.

With the help of EVS deep learning experts, you will explore style transfer literature and come up with the best model to apply a photorealistic style to a computer graphics generated image. You will have access to the required compute power and the tools needed to generate an adequate dataset. You will be responsible to design the model, implement it in PyTorch, train it and test various configurations and diverse architectures.

Emirates
FLY BETTER
JUST DO IT

Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com and o.barnich@evs.com

EVS Internships 2025-2026

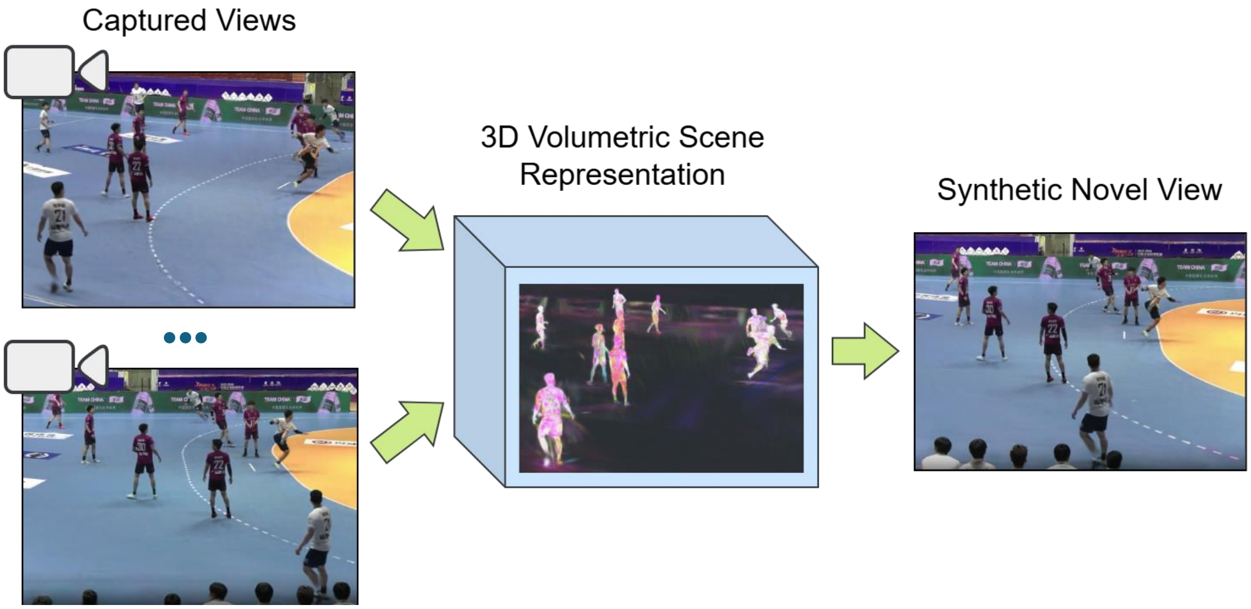
Novel View Synthesis for Immersive Sports Broadcasting

Keywords: Novel View Synthesis, Volumetric Rendering, Gaussian Splatting

Novel view synthesis has the potential to allow viewers to experience sports events from perspectives not captured by physical cameras. Recent breakthroughs in this field, such as 3D/4D Gaussian Splatting, have revolutionized volumetric rendering by offering unprecedented quality and speed compared to earlier neural radiance field (NeRF) approaches. As a teaser, check out https://zju3dv.github.io/longvolcap/ for an example of volumetric video rendering.

However, in real-world scenarios, only a limited number of cameras are often available, making it challenging for these methods to function effectively. The internship will therefore focus on exploring novel view synthesis methods capable of rendering real-time dynamic sports scenes from a limited set of cameras. You will evaluate these methods on multi-view sports datasets, paying particular attention to rendering quality, temporal consistency, and computational efficiency.

Working within our innovation department, you will collaborate with engineers and researchers to develop proof-of-concept applications that showcase the potential of these technologies for enhancing sports broadcasts. This internship offers a unique opportunity to work at the cutting edge of computer graphics and computer vision, with direct applications to next-generation broadcast technology.



Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com *and* o.barnich@evs.com

EVS Internships 2025-2026

Multimodal Embeddings for Enhanced Video Content Search and Captioning

Keywords: CLIP, Vision-Language Models, Embeddings, Video Captioning, Semantic Video Search

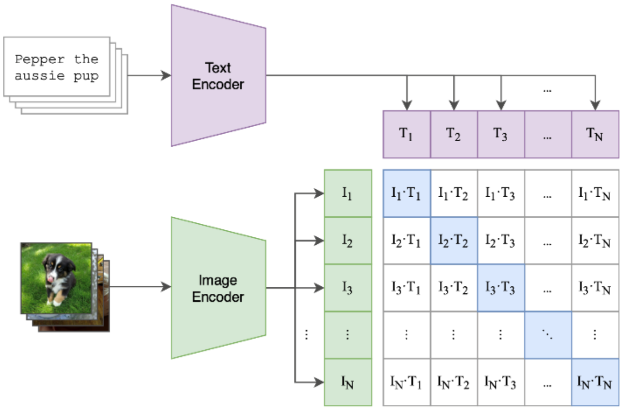
Vision-Language Models (VLMs) like CLIP (Contrastive Language-Image Pre-training) have revolutionized the way machines understand visual content by creating a shared embedding space between images and text. These models enable powerful semantic search capabilities that go beyond traditional metadata-based approaches, allowing for intuitive content discovery based on natural language descriptions.

In the broadcast industry, the ability to efficiently search, categorize, and caption vast video archives is crucial for content monetization and accessibility. Current methods often rely on manual tagging or basic computer vision techniques that fail to capture the rich semantic content of video material.

This internship focuses on leveraging CLIP-like architectures and other VLMs to develop an advanced video search and automatic captioning system specifically tailored for broadcast applications. You will explore how these models can be fine-tuned on domain-specific content to improve search accuracy and generate descriptive captions that capture the nuances of broadcast footage.

Your responsibilities will include implementing and evaluating different embedding approaches, developing efficient indexing methods for large-scale video libraries, and creating a prototype system that demonstrates improved search capabilities and automatic captioning. You will also investigate how these technologies can be integrated into existing broadcast workflows to enhance content discovery and accessibility.

This internship offers a unique opportunity to work with EVS's innovation team, applying cutting-edge AI research to solve real-world challenges in the broadcast industry. Your work will be conducted under the guidance of experienced computer vision engineers and will contribute directly to next-generation media management solutions.



	T_1	T_2	T_3	...	T_N
I_1	$I_1 \cdot T_1$	$I_1 \cdot T_2$	$I_1 \cdot T_3$	...	$I_1 \cdot T_N$
I_2	$I_2 \cdot T_1$	$I_2 \cdot T_2$	$I_2 \cdot T_3$	...	$I_2 \cdot T_N$
I_3	$I_3 \cdot T_1$	$I_3 \cdot T_2$	$I_3 \cdot T_3$	...	$I_3 \cdot T_N$
⋮	⋮	⋮	⋮	⋱	⋮
I_N	$I_N \cdot T_1$	$I_N \cdot T_2$	$I_N \cdot T_3$	...	$I_N \cdot T_N$

Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com *and* o.barnich@evs.com

EVS Internships 2025-2026

Exploring Mamba State Space Models for Efficient Video Processing in Broadcast Applications

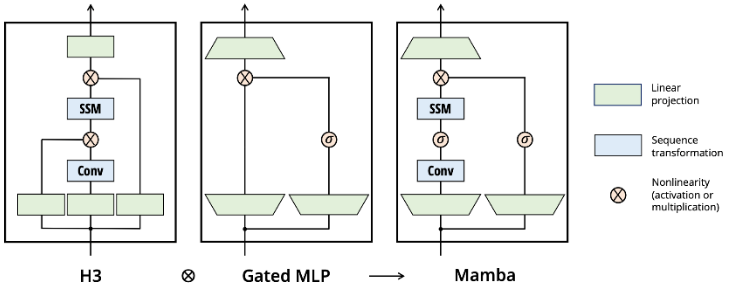
Keywords: Mamba, State Space Models, Sequence Modeling, Video Processing, Broadcast Technology

Mamba represents a breakthrough in sequence modeling through its innovative state space model (SSM) architecture. Unlike traditional Transformers, which struggle with quadratic complexity when handling long sequences, Mamba achieves linear scaling while maintaining or exceeding Transformer performance. This efficiency makes it particularly promising for video applications, where processing long temporal sequences is essential and where traditional approaches often struggle with the computational demands of processing video streams at scale, especially in live production environments.

This internship aims to explore how Mamba's architecture can be applied to video-specific tasks in broadcasting. You will investigate Mamba's potential for efficient video understanding, temporal feature extraction, and long-context processing. This could benefit applications such as automatic highlight generation, content segmentation, object tracking, super resolution, and deblurring. These tasks leverage sequence data to enhance image quality, offering substantial improvements in broadcast content delivery.

Your role will involve diving deeply into the fundamentals of the Mamba architecture, becoming an expert in its implementation for video processing tasks, and comparing its performance against current industry standards. You will develop proof-of-concept applications that demonstrate the advantages of this architecture in broadcast-specific scenarios, focusing on processing efficiency and accuracy. As an expert, you will also present the results of your work to the computer vision team, showcasing how Mamba can transform video processing workflows.

Working within EVS's innovation department, you will collaborate with a team of engineers and researchers to evaluate how Mamba can address existing limitations in video processing pipelines. This internship provides an opportunity to work at the intersection of cutting-edge AI research and practical broadcast technology applications, potentially shaping the future of video processing in the industry.



Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com and o.barnich@evs.com

EVS Internships 2025-2026

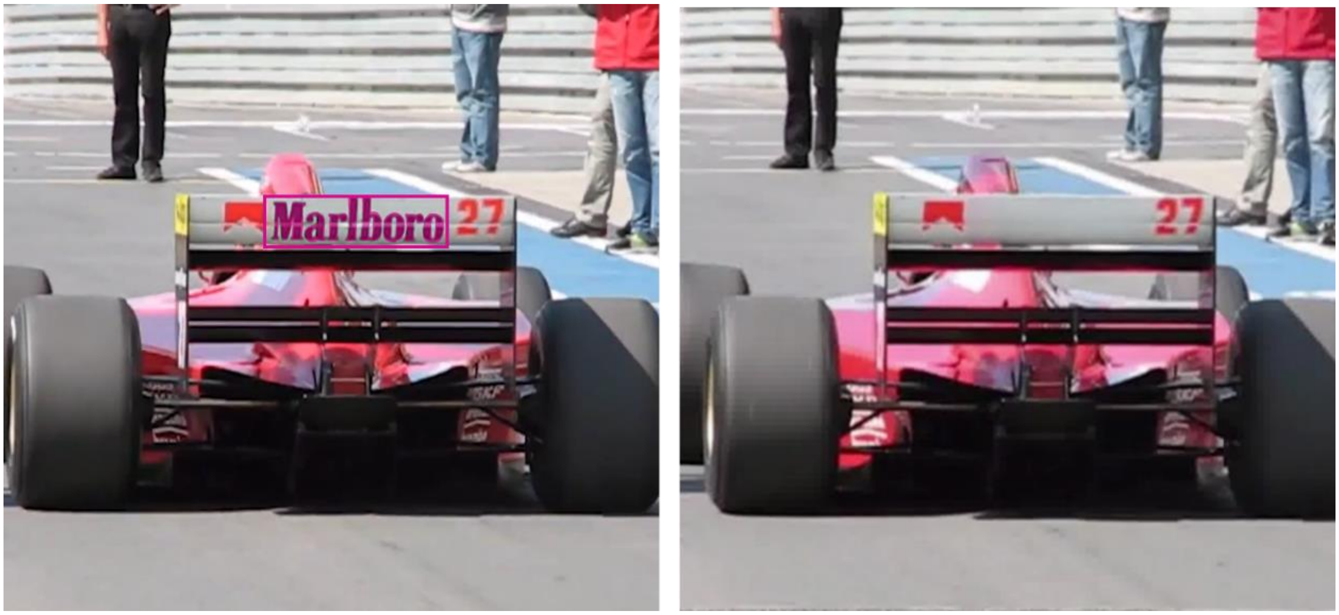
Generative models for seamless ad and object removal in videos

In some cases, broadcasters may wish to remove advertisements from video content. This is particularly the case for the re-broadcasting of certain sports content such as Formula 1, where the presence of cigarette brands is tolerated on certain platforms and prohibited on others.

Currently, the post-processing performed on these videos to comply with the standards of the broadcast platforms is carried out manually and mainly involves the use of blurring techniques. However, this blurring is often too visible and considerably degrades the perceived quality of the images, which is detrimental to the user experience.

To make this removal of ads more discreet and less invasive to the image, we want to explore the use of generative artificial intelligence models, particularly inpainting models. Inpainting consists in regenerating selected areas of an image or video in a way that is visually consistent with the surrounding context. Thanks to recent advances in diffusion models (such as Stable Diffusion), generative transformers and GANs (Generative Adversarial Networks), it is now possible to produce photorealistic results, in near-real time, on complex images or video sequences.

We already have powerful models capable of generating segmentation masks in real time that precisely locate advertising areas in images. The trainee's role will be to use these masks as input for inpainting models, with the aim of making adverts disappear in a visually consistent way. To do this, they will have to identify, train and refine several inpainting models, then evaluate them on sports content in order to determine the solution offering the best compromise between image quality and processing speed.



Interested? Send your CV and motivation letter to v.liesse@evs.com v.botta@evs.com *and* o.barnich@evs.com

EVS Internships 2025-2026

Automatic Highlights Generation for Sport Broadcasting

Keywords: content summarization, speech-to-text, video captioning, natural language processing.

In sports broadcasting, a highlight is a curated segment of an event that captures its most important, exciting, or impactful moments. These summaries typically range from quick 3–5 minute recaps to more in-depth 10–12 minute overviews, catering to audiences from casual viewers to dedicated fans. With growing demand for personalized and on-demand sports content, broadcasters are increasingly seeking efficient, scalable methods to produce multiple highlight versions tailored to various platforms and user preferences.

Recent advances in artificial intelligence—particularly in computer vision, natural language processing, and multimodal learning—have paved the way for increasingly sophisticated and high-quality automated highlight generation systems. For instance, speech-to-text technology now delivers high-accuracy performance, while large language models (LLMs) offer powerful capabilities for identifying key moments and filtering out irrelevant content.

This internship will take place at EVS headquarters in Liège, under the supervision of engineers from the Innovation team. The objective is to explore how to most effectively identify video segments that encapsulate the most relevant moments by fusing a variety of modalities, including live speech-to-text commentaries, scripted live commentaries, video captioning, and human-labeled metadata.