REAL-TIME CROWD ANALYTICS AT THE EDGE:

AI-Powered Visual Intelligence

Axelera AI explores edge-native crowd and face analytics in real time, powered by the Metis® 4-chip PCIe AI Accelerator Card and Voyager® SDK for on-device inference.

EXECUTIVE SUMMARY

Large-scale gatherings, from sold-out stadium finals to open-air music festivals and city marathons, generate enormous volumes of visual data across wide physical areas. Security teams, venue operators, and event managers are increasingly expected to monitor crowd density, track movement patterns, and respond to developing situations in real time, all while managing limited staff and strict data privacy requirements.

Computer vision is already being deployed in these environments to meaningful effect. Stadiums like Allianz Parque in Brazil and Citi Field in New York use AI-assisted systems at entry points to streamline access and reduce congestion. During the Paris 2024 Olympic Games, AI-enhanced video surveillance assisted in monitoring crowd activity across multiple venues simultaneously. In each of these cases, AI vision serves a clear purpose: helping human operators monitor complex environments, detect developing patterns, and surface concerns that would otherwise go unnoticed across hours of footage and dozens of camera feeds. The technology extends what a team can realistically see and act on, without replacing the judgment of the people responsible for those decisions.

THE WIDE-AREA VISUAL ANALYTICS CHALLENGE

Adopting these capabilities at scale, however, comes with significant practical barriers. Traditional AI inference relies on GPU-based hardware that is costly to procure and operate. Cloud-dependent architectures introduce latency that slows response times and exacerbates bandwidth issues—a real constraint for stadiums deploying dozens of cameras. They also require raw video to be transmitted off-premises, which creates friction with privacy regulations and data sovereignty requirements. Integrating new systems with existing venue infrastructure adds further complexity, and many operators find that the investment required to deploy enterprise-grade visual analytics places it out of reach for all but the largest organizations.

Monitoring high-traffic environments with hundreds of people moving through wide fields of view presents unique hurdles:

- **Computational Load:** Processing 8K video in real-time requires massive throughput that typical edge devices cannot provide.
- **Detail Preservation:** Standard systems often downscale video to fit AI models, losing the resolution needed for accurate face and object detection at a distance.
- **Privacy & Bandwidth:** Transmitting ultra-high-definition (UHD) streams to the cloud is bandwidth-heavy and introduces significant privacy compliance risks.
- **Continuity:** Maintaining a link between a "person" (body detection) and their "identity" (face detection) as they move through a crowded scene is algorithmically complex.

ARCHITECTURE FOR HIGH-RESOLUTION EDGE ANALYTICS

To illustrate how Metis and Voyager SDK can handle scenarios like these, Axelera AI developed a crowd analytics demonstration. The example processes a full 8K crowd scene in real time, performing person detection, pose estimation, face recognition, and object detection. A key technical aspect of this demonstration is its tiling strategy, configured through the Voyager SDK: users can assign smaller tiles to distant areas of the scene where fine detail matters most for inference, and larger tiles to close-up objects where that detail is already present, a more purposeful approach than uniform tiling. This is one example of how the Voyager SDK gives users control over how their

solutions are tuned.

- **AI Acceleration:** Axelera AI Metis PCIe accelerator card, featuring four AIPUs (16 total AI cores) providing up to 856 TOPS of performance (214 TOPS per chip).
- **Host Processor:** High-performance edge workstation (e.g., Intel® Core i9-12900) to manage 8K stream ingestion.
- **Software Stack**: The Voyager SDK orchestrates multiple models in parallel. It utilizes configurable tiling to divide 8K frames into smaller, high-resolution regions for independent analysis, ensuring objects in the distance are detected.
- **Models:** Simultaneous execution of Ultralytics YOLO11 (for object detection/segmentation and people detection), Ultralytics YOLOv8 (for keypoint detection), RetinaFace (for high-precision face detection), and FaceNet (for face recognition).
- **Power Efficiency:** The 4-chip Metis card operates within a 30–58W typical power draw, delivering the performance of a high-end server GPU at a fraction of the power footprint.

ACHIEVED RESULTS

Ultra-High-Resolution 8K Processing

The demonstration successfully processed a single high-resolution 8K camera at 30 frames per second (FPS). By using progressive tiling, the system analyzed the scene at full resolution, allowing for precise detection of individuals even at the far edges of wide-area venues.

People and Face Association

A distinctive capability demonstrated here is a side-by-side operator view enabled through the Voyager SDK. While the primary feed shows individuals moving through the scene in real time, a companion grid displays each tracked person's face in a fixed position, similar to a video call layout. This gives operators a stable reference point to study individuals without having to track moving subjects across a live feed.

To connect both views, the system draws a line from each person in the live feed to their corresponding position in the grid, making the association between a moving subject and their fixed portrait immediately clear. This kind of configurable operator interface illustrates how the Voyager SDK can be used to build custom viewing experiences tailored to the needs of a specific deployment.

Privacy-First Deployment

Because all inference, from 8K ingestion to face detection, runs fully on-device, no raw video ever leaves the local network. This dramatically reduces bandwidth requirements and ensures that sensitive biometric data is processed in a secure, edge-native environment.

CONCLUSION

What this example shows is that large-scale crowd analytics, combining person detection, pose estimation, face recognition, object detection and tracking (multiple AI models running in parallel) across 8K video, can operate at the edge without the cost or privacy risks associated with cloud-dependent infrastructure. The Metis AIPU provides the parallel processing capacity that high-resolution video demands, while the Voyager SDK enables the intelligent tiling strategies and model compatibility that make this level of complexity easy to process, efficient, and done in real-time. For organizations exploring how to bring visual intelligence closer to the source, the Metis platform and Voyager SDK offer the performance and flexibility to deploy production ready solutions today.

ABOUT AXELERA:

Axelera AI is the leading provider of purpose-built AI hardware acceleration technology for AI inference, including computer vision and generative AI applications. Its first-generation product is the game-changing Metis™ AI platform – a holistic hardware and software solution for Edge AI inference which delivers the world's highest performance and power-efficiency at a fraction of the cost of alternative solutions. Headquartered in the AI Innovation Center of the High Tech Campus in Eindhoven, The Netherlands, Axelera AI has R&D offices in Belgium, Switzerland, Italy and the UK, with more than 200 employees in 18 countries. Its team of experts in AI software and hardware hail from top AI firms and Fortune 500 companies.

Visit us at AXELERA.AI