intel. + iSIZE

iSIZE BitClear: Deep Perceptual Denoising and Upscaling with Intel® Advanced Matrix Extensions

iSIZE BitClear takes advantage of Intel® Advanced Matrix Extensions (Intel® AMX) on 4th Gen Intel® Xeon® Scalable processors to process video on a real-time basis, without requiring an external accelerator such as a GPU. Performance measured by iSIZE shows that the high per-core throughput and built-in Intel AMX accelerator delivers up to 2X the performance of the 3rd Gen Intel® Xeon® Scalable processor[1] and requires significantly fewer processor cores than the latest-generation competitor CPU.[2]

Video content continues to grow dramatically in terms of sheer volume, with the most dramatic growth in material produced informally or under less-than-ideal conditions. For example, ubiquitous smartphones have made everyone a content producer, and as a result, more than 500 hours of video are uploaded to YouTube every *minute*.[3] Likewise, machine generated video has become more ubiquitous. The global market for security and surveillance cameras is expected to grow by a CAGR of 18.7% through 2030,[4] and the global market for drone cameras is expected to grow over that same period at a CAGR of 26%.[5]

The suboptimal circumstances under which these videos are captured, stored and transmitted often drive down quality. A range of factors such as dirty lenses, poor focus and camera shakiness from user movement or drone vibrations can compromise content. Capturing video using low-power and low-cost solutions or transmitting it over low-bandwidth or low-quality wireless links can introduce additional compression artifacts. Even when videos are captured and transmitted in ideal conditions, quality can still be degraded due to storage or transmission limits on end devices causing only low-resolution versions to be kept or uploaded, or because of re-encoding as they are re-shared and re-uploaded.

BITCLEAR — Deep Perceptual Denoising and Upscaling

- Video from smartphones, security cameras, drones and older professional cameras often has poor quality due to poor conditions or encoding artifacts

- BitClear from iSIZE solves these problems through AI-based neural network technology that can run in real-time on Intel Xeon Scalable processors

- 4th Gen Intel Xeon scalable processors with Intel AMX provide further performance improvements: benchmarks by iSIZE show up to 2X over 3rd Gen Intel Xeon Scalable CPUs[1] and requirements for significantly fewer processor cores than the latest-generation competitor CPU[2]

2X

performance improvement over 3rd Gen Intel® Xeon® Scalable processors

In addition to quality concerns, video is often uploaded from smartphones at low resolutions such as 360p or 720p, which are insufficient for the standards expected by modern streaming and social media services. In particular, such video typically looks too small on full-size displays, and its quality can drop significantly if it is enlarged using the built-in consumer-grade upscaling available in mainstream displays.

Beyond user-generated content, even professional content can benefit from quality improvements. For example, due to infrequent upgrade cycles of cameras for live events (concerts, stadiums, etc.), live video may not be captured in the 4K format and contribution encoding quality needed for the latest OTT streaming or broadcast standards.

AI-based denoising and upscaling solutions are the standard approach to overcoming these limitations, providing for sharper edges and textures and increasing image resolution intelligently. The workloads associated with these processes are processing-intensive, and running them on CPUs unaided is often insufficient. As a result, it is common to deploy discrete GPUs or other special-purpose hardware accelerators for this function.

Because such hardware tends to be expensive and power-hungry, it can add significantly to server equipment and operating costs, while also not being applicable to general-purpose workloads. Moreover, it is challenging to provide the high throughput and low latency required, especially for real-time processing of live video such as surveillance feeds from security cameras or drones.

iSIZE BitClear, running on Intel Xeon Scalable processors, is an AI-based solution for both denoising and upscaling that can run in real-time without GPUs or special-purpose accelerators. It can provide real-time performance to handle live video or efficient batch processing of multiple simultaneous video uploads.

BitClear: CPU-Based Enhancement for Video Quality and Resolution

iSIZE leveraged its deep learning specialty in video delivery to create BitClear, an AI-based neural network technology focused on denoising and upscaling of compressed video to significantly improve its perceptual quality. BitClear enables customers to share, host or otherwise provide higher quality media content that better engages the targeted audience. The solution achieves this goal by using pre-trained AI models to remove compression artifacts and optionally upscale the video content, as illustrated in Figure 1.

Unlike competing approaches, BitClear operates in real-time and at scale without manual tuning, as it is trained to disentangle the source and noise manifolds and recover video details without changing the aesthetics of the decoded video.

Intel AMX, an accelerator built into 4th Gen Intel Xeon Scalable processors, increases performance by up to 2x compared to 3rd Gen Intel Xeon Scalable processors,[1] and requires significantly fewer processor cores than the latest-generation competitor CPU.[2] The number of real-time streams that can be handled by the combination of BitClear and 4th Gen Intel Xeon Scalable processors is better than what can be done on high-end GPUs, allowing customers to leverage widely available CPUs in their own infrastructure or in the public cloud, without the fixed cost of dedicated GPU infrastructure.

How BitClear Works

Depending on customer preferences or requirements, the solution can be deployed on-premise or in the cloud. When a low-quality or heavily compressed media file is received, the BitClear solution removes compression artifacts after decoding to restore video quality, as shown in Figure 2. The content can then be sent for further processing or
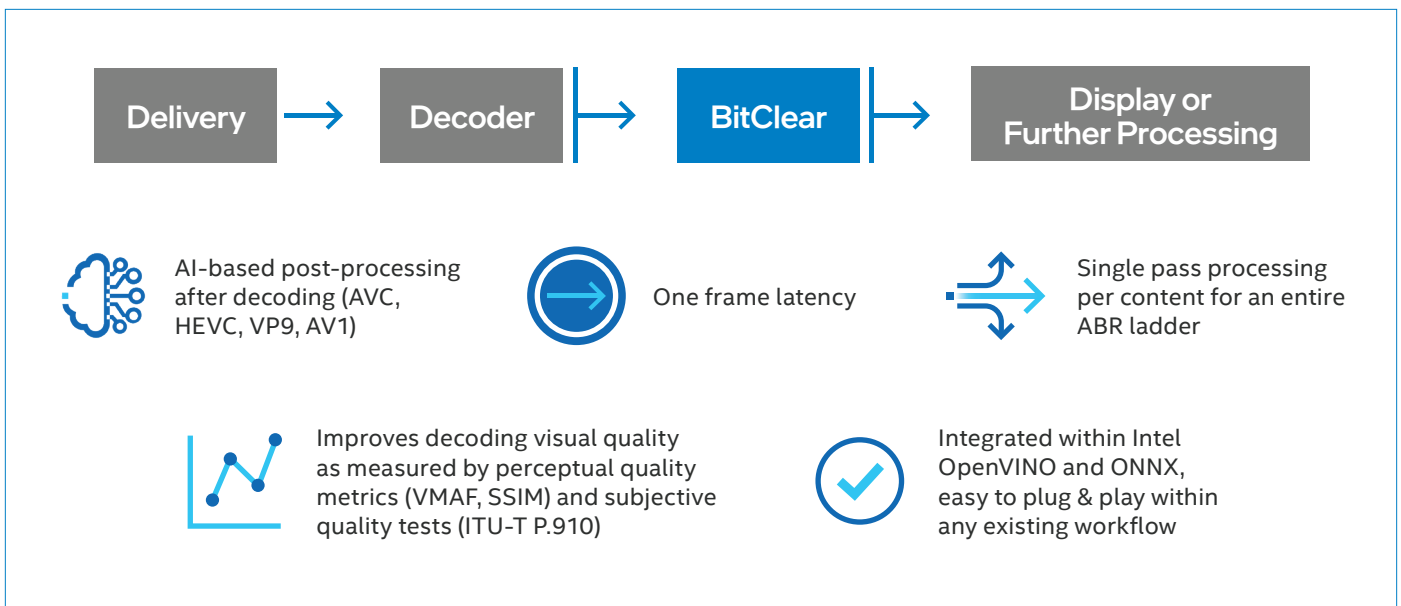

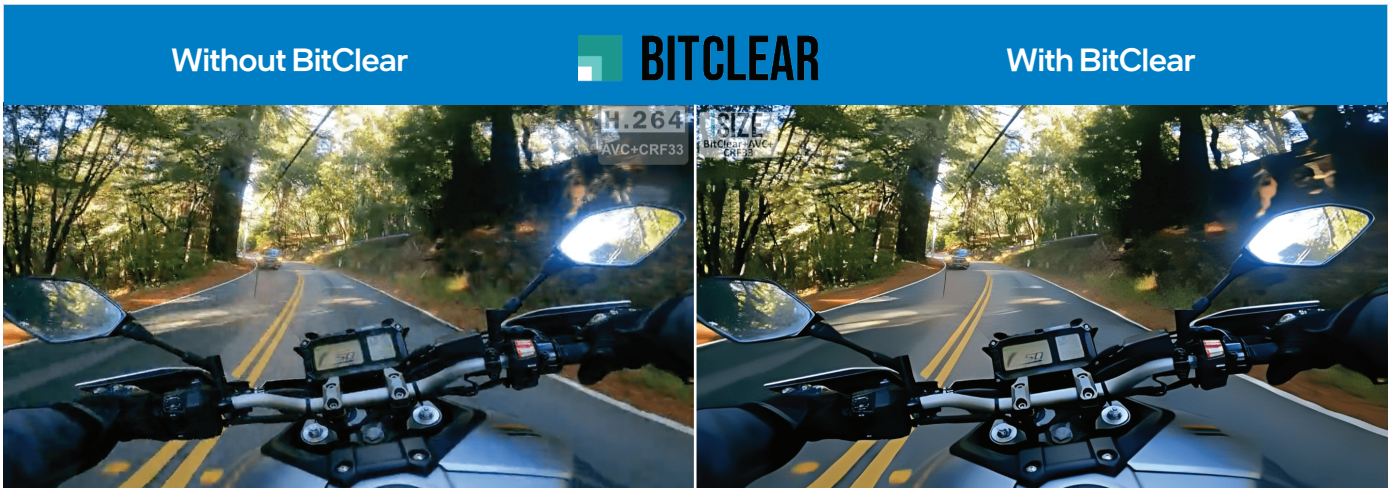**Figure 1.** iSIZE's BitClear pipeline for AI-based removal of compression artifacts.

Figure 2. Visual examples of quality improvement using BitClear.

transcoding, or it can be stored, streamed or shared to the customer's chosen destination. BitClear can be applied as a post-processor to bitstreams received by any device.

Rigorous ITU-T P.910 ACR testing, as well as objective quality metrics (VMAF), show that BitClear offers significant quality improvement. In addition, the solution offers the features and capabilities shown in Figure 3.

Accelerated Deep Learning with Built-in Intel AMX

Intel AMX is an accelerator built into 4th Gen Intel Xeon Scalable processors that accelerates AI for the BitClear solution by speeding up the tensor processing at the heart of deep learning algorithms. Workloads across data centers and operator networks use Intel AMX to accelerate both training and inference for implementations such as natural language processing, recommendation systems and image recognition.

Real-Time Processing

- Enables multiple input video assets in real-time[6]
- Upscales video in real-time to allow single-pass processing with single-frame latency and high performance on Intel® Xeon® processors or other CPUs

Codec Independence

- Processes any highly compressed content and produces a higher-quality output that improves the value of the asset
- Operates on all types of codecs and upholds all necessary standard compliance

Superior Quality

- Revives video back to the maximum possible resolution quality without affecting the artistic intent of the original creator
- Enables video denoising speeds up to 10 times faster than competing state-of-the-art denoising technology

Optimized Scalability

- Scales across high volumes of content with neural network architecture
- Satisfies standard perceptual quality metrics and visual quality as assessed by human viewers under controlled test conditions
- Deploys either in the cloud or on-premise

Figure 3. BitClear features and capabilities.

Because the accelerator is built into the processor, no external hardware is needed, and the solution avoids the latency (and energy footprint) of passing data across the PCIe bus to an external device. This low-latency operation is crucial for the real-time operation of BitClear on live video feeds. Intel AMX can help BitClear implementations achieve better performance as well as CapEx and OpEx savings:

- **Performance**. Purpose-built accelerators deliver significant gains in throughput for their targeted workloads.

- **Equipment costs**. Because the accelerators are built into 4th Gen Intel Xeon Scalable processors, they do not require a separate equipment investment.

- **Operating costs**. By reducing the need for additional cores to be added to equipment racks, built-in accelerators may provide significant energy savings.

For more on Intel AMX, visit https://www.intel.com/content/www/us/en/products/docs/accelerator-engines/advanced-matrix-extensions.html.

Performance Comparison Across Hardware Platforms

To assess workload handling of various hardware for real-time video processing with iSIZE BitClear, performance teams from iSIZE and Intel compared key performance indicators (KPIs) generated on various platforms. Table 1 presents measurements by iSIZE that shows the frames per second at various resolutions for 56-core 4th Gen Intel Xeon Scalable processors, with comparative results from the predecessor Intel Xeon platform and a dedicated GPU.

3rd Gen Intel Xeon Scalable processors, which incorporate Intel Deep Learning Boost hardware acceleration for deep learning, can support up to 47 fps for 720p60 and 21 fps for 1080p60 resolution. 56-core 4th Gen Intel Xeon Scalable processors with Intel AMX can support up to 92 and 42 fps, respectively, for the same video formats, with support for real-time streams up to 720p60 and 1080p30. The same 4th Gen Intel Xeon Scalable processor can support up to 518 fps of 360p60 video, which will process a single video at much faster than real-time, or six to eight separate 360p60 streams in real-time. This is roughly 2x the performance of the 3rd Gen Intel Xeon Scalable processor, and a significant improvement over a dedicated NVIDIA Quadro RTX A6000 GPU for the case without upscaling. Given that parts of the iSIZE BitClear upscaling components are implemented in fp32 operations for CPUs (and hence cannot yet leverage AMX acceleration), further speed improvements are expected for CPUs when fully-int8 based upscaling will be supported.

Table 2 shows data from iSIZE with the number of CPU cores needed for BitClear to process a real-time video stream. For public cloud or virtualized infrastructure, 4th Gen Intel Xeon Scalable processors with Intel AMX significantly reduce the number of CPU cores needed to process a single real-time video stream (at 30 or 60 fps). The number of cores needed is approximately half what is required on 3rd Gen Intel Xeon Scalable processors.

Table 1. Frames per second per processor for real-time BitClear with and without 2X upscaling. Higher numbers are better. Data provided by iSIZE.

Input Resolution	Upscaled Resolution	Intel® Xeon® 8380 Processor (40 cores @ 2.3 GHz with Intel DL Boost)[7]		Intel Xeon 8480+ Processor (56 cores @ 2.0 GHz with Intel AMX)[8]		NVIDIA Quadro RTX A6000 GPU[9]	
		Without Upscaling	With Upscaling	Without Upscaling	With Upscaling	Without Upscaling	With Upscaling
360p60	720p60 (HD)	249	239	518	392	317	354
480p60	960p60	103	102	208	153	135	155
540p60	1080p60 (FuHD)	82	83	163	120	111	120
720p60	1440p60 (QHD)	47	46	92	67	63	67
1080p30	2160p30 (UHD)	21	20	42	31	28	32

Table 2. Number of CPU cores for BitClear to process a single real-time video stream, with and without 2X upscaling. Lower numbers are better. When hardware cannot support the target fps (60 or 30), the actual fps achieved is shown in parentheses. Data provided by iSIZE.

Input Resolution	Upscaled Resolution	Intel® Xeon® 8380 Processor (40 cores @ 2.3 GHz with Intel DL Boost)[7]		Intel Xeon 8480+ Processor (56 cores @ 2.0 GHz with Intel AMX)[8]	
		Without Upscaling	With Upscaling	Without Upscaling	With Upscaling
360p60	720p60 (HD)	7	7	3	4
480p60	960p60	16	17	8	11
540p60	1080p60 (FHD)	22	24	10	14
720p60	1440p60 (QHD)	40 (47 fps)	40 (46 fps)	20	35
1080p30	2160p30 (UHD)	40 (21 fps)	40 (20 fps)	22	46

Conclusion

In a world of fast-multiplying video from low-quality sources, iSIZE BitClear plays an instrumental role with AI-powered denoising and upscaling. Running on 4th Gen Intel Xeon processors with the built-in Intel AMX hardware accelerator for deep learning, this solution can perform these functions in real time on the CPU alone, without add-on discrete GPUs or other hardware accelerators, delivering an outstanding combination of performance and cost-effectiveness.

Learn More

www.intel.com/xeon/scalable

More Information

Intel Network Builders – iSIZE: https://networkbuilders.intel.com/ecosystem/isize-technologies

iSIZE Solutions: https://www.isize.co/oursolutions/

intel. + **iSIZE**

[1] Based on comparison of data provided by iSIZE and shown in Tables 1 and 2. See footnotes for Tables 1 and 2 for testing and configuration details.

[2] Based on comparison of data provided by iSIZE in Table 2 and published at https://www.isize.co/boosting-sustainability-in-video-streaming-amd. For example, as shown in Table 2, the 4th Gen Intel Xeon processor can process (denoising without upscaling) a single real-time video stream of 540p60 using 10 processor cores, while the data in Table 4 in the iSIZE post shows that the AMD 9654 requires 30 cores for the same workload. See footnotes for Table 2 for testing and configuration details for the 4th Gen Intel Xeon processor data.

[3] Statista, February 2020. "Hours of video uploaded to YouTube every minute as of February 2020." https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

[4] Emergen Research, August 2022. "Security Cameras Market, By System (Analog System, IP based System), By Camera Resolution (HD, Full HD), By Type (Indoor, Outdoor) By Application (City Infrastructure, Commercial), and By Region Forecast to 2030." https://www.emergenresearch.com/industry-report/security-cameras-market.

[5] The Brainy Insights, March 2022. Drone Camera Market Size by Type (SD Camera and HD Camera), Application (Surveillance, Photography & Videography, and Thermal Imaging), End User (Commercial Homeland security and Military), Regions (North America, Europe, South America, Asia Pacific and The Middle East and Africa), Global Industry Analysis, Share, Growth, Trends, and Forecast 2022 to 2030." https://www.thebrainyinsights.com/report/drone-camera-market-12672.

[6] iSIZE, September 2022. "iSIZE BitClear: Deep Perceptual Denoising and Upscaling for Video Delivery Systems." https://www.isize.co/wp-content/uploads/2022/09/iSIZE-Datasheet_BitClear_v20220905.pdf.

[7] Tested by iSIZE on Dec. 11, 2022. System configuration details: 3rd Gen Intel Xeon 8380 processor, 256GB system memory (16x16GB Hynix HMA82GR7CJR8N-XN rank2), SanDisk SSD Plus (480GB), Intel Corp. BIOS ver. SE5C6200.86B.0022.D64.2105220049, CPU microcode ver. 0xd000375, Intel Turbo Boost Technology Enabled, Hyperthreading enabled (but not used via node locking to physical cores via OpenVINO commands), Ubuntu ver. 22.04.1 LTS with 5.17.0-1021-oem kernel patches.

[8] Tested by iSIZE on Dec. 11, 2022. System configuration details: 4th Gen Intel Xeon 8480+ processor, 256GB system memory (16x16GB Samsung M321R2GA3BB0-CQKVG rank1), SanDisk SDSSDA-1 (1TB), AMI BIOS ver. 3A06.uh, CPU microcode ver. 0x2b000081, Intel Turbo Boost Technology Enabled, Hyperthreading enabled (but not used via node locking to physical cores via OpenVINO commands), Ubuntu ver. 22.04.1 LTS with 5.17.0-1021-oem kernel.

[9] Tested by iSIZE on Dec. 13, 2022. System configuration details: 2nd Gen Intel Xeon 6226R processor, 384GB system memory (12x32GB DDR4 3200 MT/s), 1x 7.3T Samsung_SSD_870, 1x 50G VIRTUAL-DISK, 1x 1.8T INTEL SSDPE2KX020T8, Supermicro BIOS ver. SYS-4029GP-TRT2 3.4, CPU microcode ver. 0x5003302, Intel Turbo Boost Technology Enabled, Hyperthreading enabled, Ubuntu ver. 20.04.5 LTS with 5.4.0-124-generic kernel, NVIDIA Quadro RTX A600 GPU with 256 GB memory.