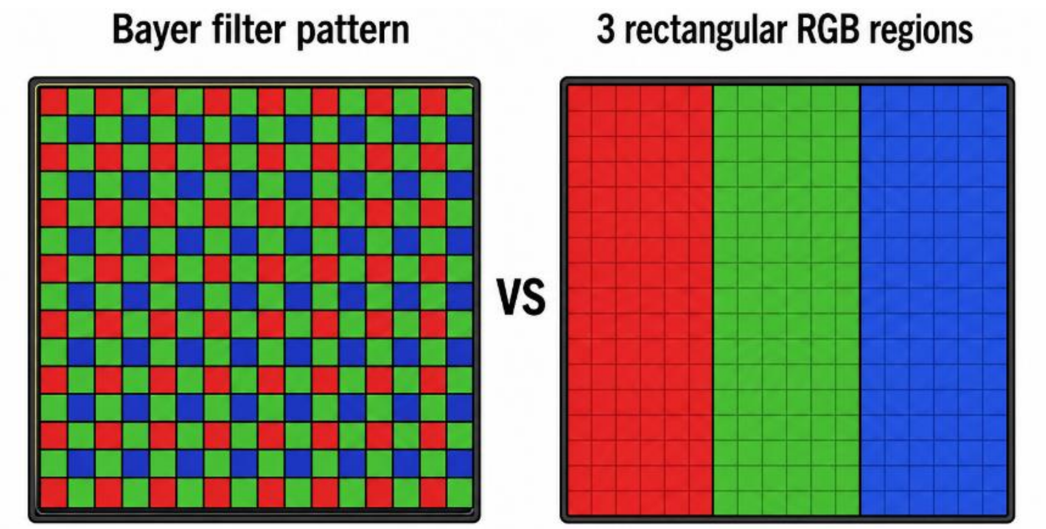
AI Algorithms for Metalens Cameras

By Chris Chinnock, Insight Media

In collaboration with MetaOptics Technologies

June 2026



Contents

Contents 2
Introduction 3
The Three-Color Metalens: Strict Color Separation by Design 3
Classical Image Processing 5
DNN Restoration Model for Metalens Imaging 7
Performance 9
  Optical performance 9
  Processing Stack and Speed 12
Applications 14
  Touchless Fingerprint Recognition 14
  Gesture Recognition 15
From Prototype to Product 15
Contact 16

Introduction

Camera modules are everywhere — phones, cars, medical instruments, biometric readers, and smart glasses. In all of them the optical front-end is the dominant cost and bulk: a stack of curved glass elements is required to deliver a sharp image to a sensor much smaller than the lens itself.

Metalenses change this picture. A metalens is a sub-wavelength nanostructured pattern on a thin glass substrate, fabricated wafer-scale by standard semiconductor DUV immersion lithography — the same process that produces logic chips. The metalens is fabricated atop a glass substrate with structures that are only a few hundred nanometers thick. This can enable radical simplification of the optical module with yield and cost scaling excellently in mass-production.

Conventional refractive lens stacks reduce aberrations by introducing different lens materials and complex lens shapes increasing complexity and cost. On the other hand, metalenses suffer from angular optical aberrations, chromatic distortions, peripheral blur, and illumination sensitivity which degrade perceptual quality. MetaOptics Technologies seeks to close this gap via software.

The optical design is intentionally minimal — one metalens for monochrome imaging, or three adjacent metalenses with per-channel filters on a single monochrome sensor. Image-processing pipelines built and tuned in MetaOptics Technologies' proprietary AI development software enhance the image in ways that can be tailored to specific camera sensors and applications. Such algorithms can be exported to the customer's deployment platform thus forming a bridge to the eventual on-chip ISP.

While image processing software can also be used with conventional refractive optical lens stacks, we believe the combination of a metalens plus image processing software will offer a better option in terms of performance, cost and volume for certain applications. In particular, we believe the biometric sensor application market is perfectly suited to adopt this approach.

The Three-Color Metalens: Strict Color Separation by Design

A conventional color camera carries a Bayer filter array on top of its sensor: each full color pixel is composed of one red and blue sub-pixels and two green sub-pixels. The full color image is created with a demosaicing algorithm. A significant part of the color information is therefore inferred rather than measured, and the resulting image carries color-aliasing artifacts on every fine edge.

MetaOptics Technologies' color metalens replaces this with three adjacent metalenses on a single glass substrate, each one projecting a full Field-of-View (FOV) image of the scene onto its own region of a monochrome sensor (Figure 1). Uder each region, red, green and blue color filters create these image components. The sensor sees no Bayer pattern, so no demosaicing is required and there is minimal channel-to-channel crosstalk. Each metalens is designed for its own spectral band so every pixel of a channel is a direct measurement of photons in that band.

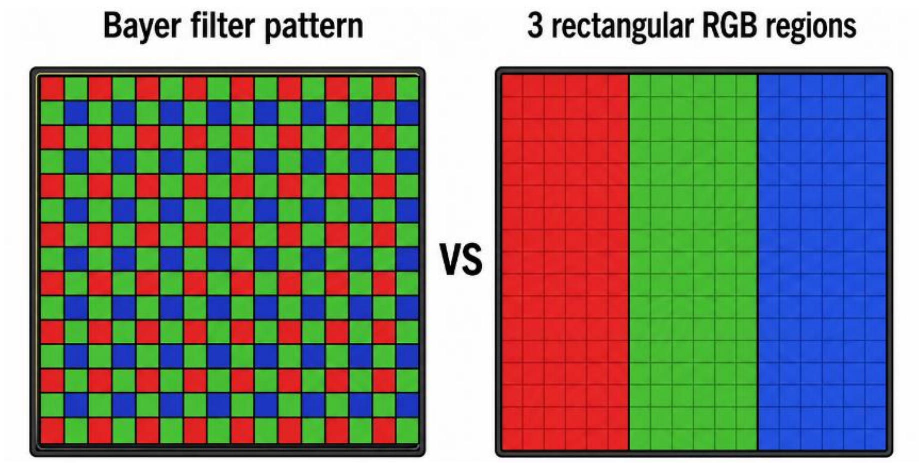


Figure 1: Standard Bayer filter pattern versus MetaOptics RGB camera filter pattern. The Bayer filter pattern has an interleaved RGB pixel mosaic whereas MetaOptics splits its sensor area into dedicated R, G, and B zones.

The benefits of this strict color separation propagate through the entire pipeline:

- **No demosaicing processing.** There is no demosaicing needed saving processing and resulting in cleaner edges.
- **Improved resolution -** A Bayer filter uses 4 sub-pixels to create a color pixel, whereas the MetaOptics RGB array uses only 3 sub-pixels to create a full-color image.
- **Minimal color crosstalk.** Narrow-band per-channel filtering produces saturated, faithful color and clean edges, especially in challenging high-contrast scenes.
- **Improved Fill Factor.** The metalens shape can match the sensor footprint, so almost no captured light is wasted. Conventional optics are round and bigger than a rectangular or square sensor profile.
- **Optical depth as a free signal.** Because the three lenses are laterally displaced on the substrate, every world point at finite distance projects to slightly different pixel positions in each channel — a passive depth cue used later for three-dimensional applications.
- **Extreme compactness.** The metalens itself is a few hundreds of nanometers thick, plus the glass substrate thickness, opening a roadmap to a module thickness that a multi-element refractive stack cannot reach.

Modern edge device image processing stacks for conventional camera modules leverage on cutting edge hardware, execution engines, optimization techniques and software platforms (see Fig. 9). A stack tailored to and optimized for metalens imaging can greatly boost its imaging performance. A critical step is hardware Image Signal Processing (ISP) development to efficiently facilitate basic processing on chip. Without ISP the color image must be assembled in software by aligning and combining the three sub-images. That image-fusion step is the entry point to the rest of the processing pipeline.

Classical Image Processing

Image processing and enhancement are core components of modern camera systems and are widely used in smartphones, consumer electronics, industrial vision systems and embedded optical modules. In compact cameras, the final image quality is not determined by the optics alone. Instead, it results from the combined performance of the optical design, sensor, illumination conditions and image-processing pipeline. This is particularly relevant for metalens camera modules, where the optical stack can be made very thin, but the captured image may require dedicated processing to compensate for system-specific effects such as reduced brightness, sensor noise, color-channel misalignment, blur or non-uniform image quality across the field.

For RGB metalens camera modules, image fusion is typically the first processing stage. The raw monochrome sensor frame contains separate blue, green, and red monochromatic images of the same scene, captured through different sub-apertures, as shown in Fig. 2. These sub-images must be spatially aligned and combined into a single-color image with high quality as it is the basis for all subsequent processing stages.

Figure 2: Rectangular metalens image fusion. Left: Raw monochrome sensor frame from MetaOptics RGB metalens camera containing blue, green, and red monochromatic images of the same scene before fusion to a raw color image. Right: after the fusion stage where the three sub-apertures are aligned and combined into the raw color image.

After fusion, classical image-processing methods include white balance, contrast equalisation, morphological filters, Gabor filters, edge-aware smoothing and sharpening. Such methods can be useful for metalens image enhancement or image filtering for different applications. Low-light scenes require particular care. Increasing the sensor gain or applying a uniform digital gain can brighten the image, but it can also saturate already bright regions and amplify noise in darker areas. A more selective brightening approach can be used to lift dark regions more strongly, adjust mid-tones moderately and preserve highlights, as illustrated in Fig. 3. In this approach, multiple virtual exposures are blended in a perceptual color space using multi-scale image pyramids. This allows well-exposed information to be selected locally across the image, improving visibility in dark regions while maintaining natural contrast and avoiding excessive highlight clipping.

Figure 3: Image enhancement on a RGB metalens camera. Left: An example raw image after image fusion. Middle: Classical processing — basic fusion, brightness, and white balance. Right: After deep image-restoration stage — edge sharpness restored across the full field, uniform color rendition from center to corner. Middle and Right image are taken from a slightly different perspective compared to the example raw image.

Figure 3 shows an example processing sequence for an RGB metalens camera. The left image shows an example raw image after image fusion. The middle image shows the result after classical processing, including basic fusion, brightness correction, and white balance. The right image shows the result after an additional deep image-restoration stage, where edge sharpness is restored across the full field and the color rendition becomes more uniform from the center to the corner. The middle and right images are taken from a slightly different perspective compared to the example raw image.

Classical processing steps can provide a flexible foundation for improving metalens camera images before comprehensive more advanced AI-based enhancement is applied. At MetaOptics Technologies, such processing pipelines are developed, tested and optimized in MOTviewer, the company's in-house software environment for AI-based image processing development, camera evaluation and prototype demonstrations. MOTviewer allows conventional processing nodes to be combined with AI-based modules in application-specific pipelines and supports deployment-oriented development for target devices (see MetaOptics AI Vertical at www.metaoptics.sg). Building on this classical image-processing foundation, the next section discusses AI-based restoration and enhancement methods developed for metalens camera modules.

DNN Restoration Model for Metalens Imaging

Beyond conventional processing, MetaOptics also develops deep-neural-network (DNN) restoration models for metalens imaging (DRMI). These AI models are trained to improve image quality by learning the relationship between metalens captures and higher-quality reference images. The idea is based on the design knowledge of the metalens point-spread function, which can vary across the field of view. A neural network can implicitly learn how to compensate for field-dependent blur, color non-uniformity and other image degradations. The result is a restored image with improved sharpness, more uniform color rendition and enhanced visual quality in a single processing pass.

Image degradation can be more pronounced in peripheral image regions and are further compounded under non-uniform or low illumination, which frequently occurs in practical settings. Existing restoration models still suffer from two major limitations.

- First, they struggle to recover details in the outer regions of the image due to spatially varying degradations caused by angular aberrations. These regions are typically under-optimized during training, as most models implicitly focus on the brighter and more informative image center.
- Second, their performance drops significantly when tested under unseen illumination levels, because they are usually trained under fixed lighting conditions and fail to generalize across varying illumination profiles.

These two challenges—spatial degradation and illumination sensitivity—are the main bottlenecks impeding robust deployment of metalens-based systems.

A key enabler of the framework is the IlluMeta dataset—a large-scale, real-world metalens image dataset captured under diverse illumination conditions. Existing datasets either lack real illumination variation or are synthetic in nature making it difficult to study illumination-aware restoration in realistic scenarios. In contrast, IlluMeta contains over 25,000 real-world metalens images, captured across 13 distinct illumination levels, and includes both daytime and nighttime scenes. This dataset enables the systematic evaluation of models under varying lighting conditions, encourages research into generalizable restoration techniques, and provides rich spatial degradation patterns for studying position-dependent image quality.

The two-pronged restoration framework explicitly targets the above challenges. The first is a Spatial-Aware Attention Scheduler that encourages the model to pay greater attention to peripheral regions during training. This design is motivated by the observation that angular aberrations introduce stronger degradations away from the image center, and that treating all regions equally causes the model to neglect these harder regions. By weighting the reconstruction loss spatially, our method facilitates better structural fidelity across the entire image. The result is a measurable improvement in peripheral image quality (Fig. 4).

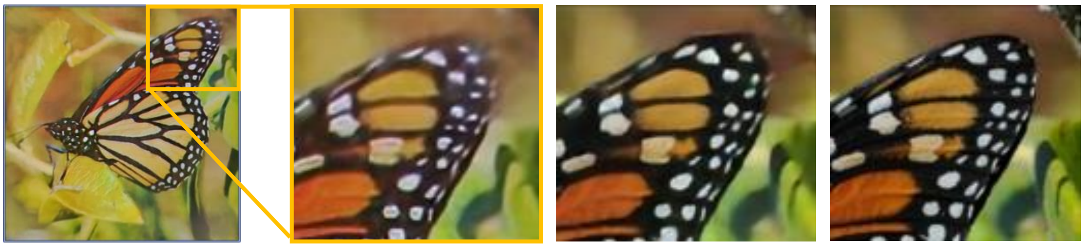


Figure 4: Left: Image with zoom-in showing peripheral image distortion without the Spatial-Aware Attention Scheduler; center: Image restoration with the Spatial-Aware Attention Scheduler; right: Ground Truth image.

Fig. 5 shows some experimental results of the image processing over the illumination-based dataset using PSNR as a metric[1]. The algorithm was used in conjunction with four other established image processing algorithms. The data show results signal integrity is very similar for illumination level like the baseline illumination level (12) but are much improved at the extremes of illumination.

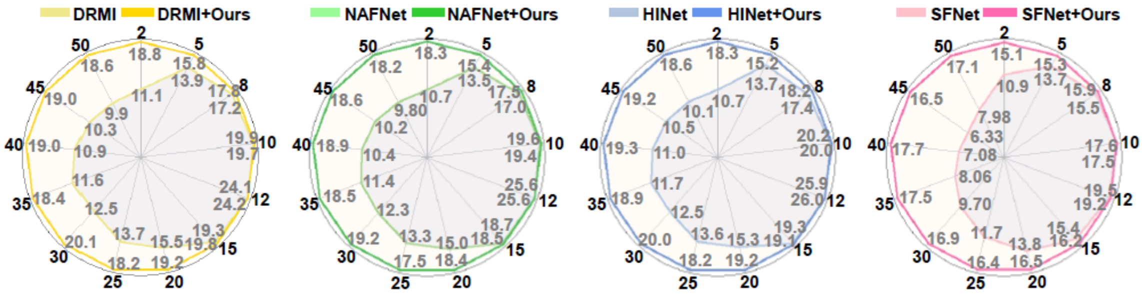


Figure 5: Comparison of PSNR (dB) between baseline image processing models (DRMI; NAFNet; HINet: SFNet) and their improved counterparts across all illumination levels (ILs). The restoration models are trained on IL 12, the RL agents are trained on a subset of ILs {2, 10, 20, 30, 40, 50}.

This is achieved by implementing a reinforcement learning (RL)-based Retraining-Free Illumination Adapter that adaptively modifies the illumination profile of the input image at inference time. Rather than retraining the restoration model for each lighting condition—which is both computationally expensive and hard to scale—a lightweight RL agent learns to adjust image illumination on-the-fly to align with the pretrained model's preferences. The agent runs and adapts the input illumination profile on the fly, guided by a quality-of-restoration reward signal. The restoration network itself is left untouched. The system therefore generalises to lighting conditions it has never seen, with no additional data collection and no retraining (Figure 6).

[1] Towards Illumination-Aware Restoration of Metalens-Captured Images:
A New Dataset and a Strong Baseline - The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

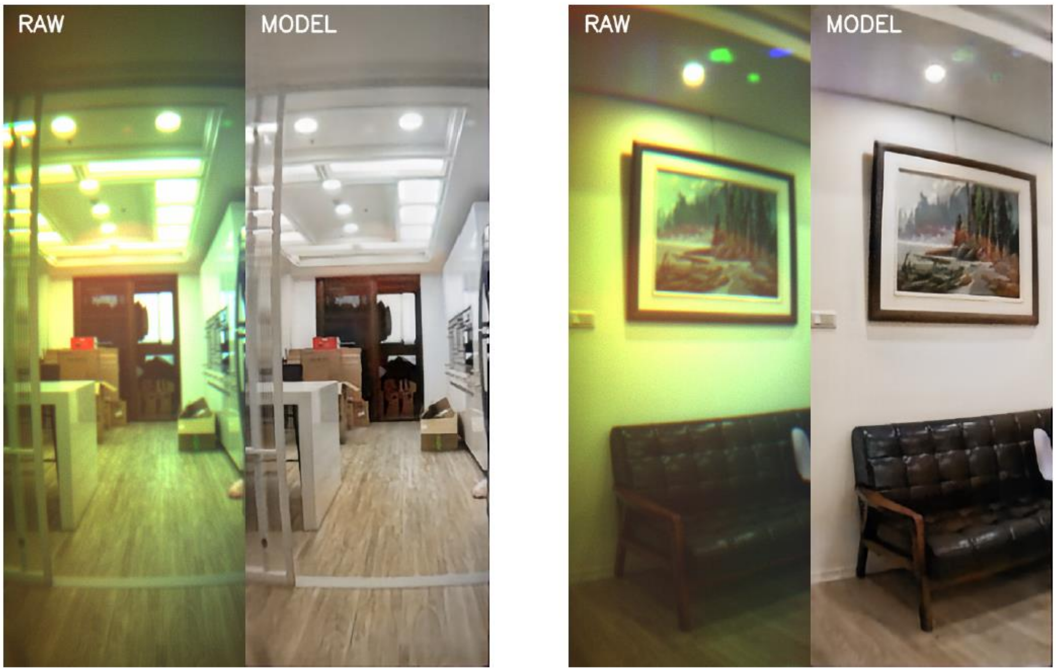


Figure 6: MetaOptics latest AI model for rectangular metalens camera image enhancement.

Performance

The preceding sections described the metalens optics and the processing algorithms separately. This section brings the two together and asks the practical question: what image quality does a single-layer RGB metalens camera deliver today, and at what speed? We first quantify optical performance on a standard resolution chart, benchmarking the basic-processed and AI-enhanced metalens output against a modern smartphone camera, and then describe the processing stack that produces these results in real time.

The comparison is deliberately demanding. The smartphone pairs a multi-element refractive lens stack with dedicated hardware ISP and years of tuned computational photography, while the metalens module forms its image with a single nanostructured thin film on a glass substrate, a monochrome sensor and a software pipeline. We present the results honestly, including where the smartphone stays ahead, because the gap that remains — and the headroom left to close it — is itself the message.

Optical performance

It is tough for single-layer RGB metalens imaging to be compared to a modern smartphone camera module. In the following we perform a quality benchmark using a Vivo V40 Lite with a Sony IMX882 (LYTIA 600-class) sensor that has 50 MP, 0.8 µm native pixels and approximately 1/1.95″ size. This sensor produces a 2×2-binned 12.5 MP output, 4080 × 3060 pixels, with an effective pixel pitch of approximately 1.6 µm, captured at a 52 mm-equivalent field of view.

Processing: The phone applies its full computational-photography pipeline. The embedded maker note confirms that HDR (HdrStatus: auto), AI scene detection, and sharpening were active. The high contrast and slight edge overshoot are products of this pipeline. These effects are distributed across the sensor, including capture and on-sensor HDR modes; the SoC image-signal processor, including demosaicing, denoising, local tone mapping, and sharpening; and the vendor's camera app, including multi-frame HDR fusion and the tuning that determines the strength of these processing steps. A bare sensor RAW would look much closer to the basic frames. In this sense, the comparison is one enhancement pipeline against another, not against an unprocessed ground truth.

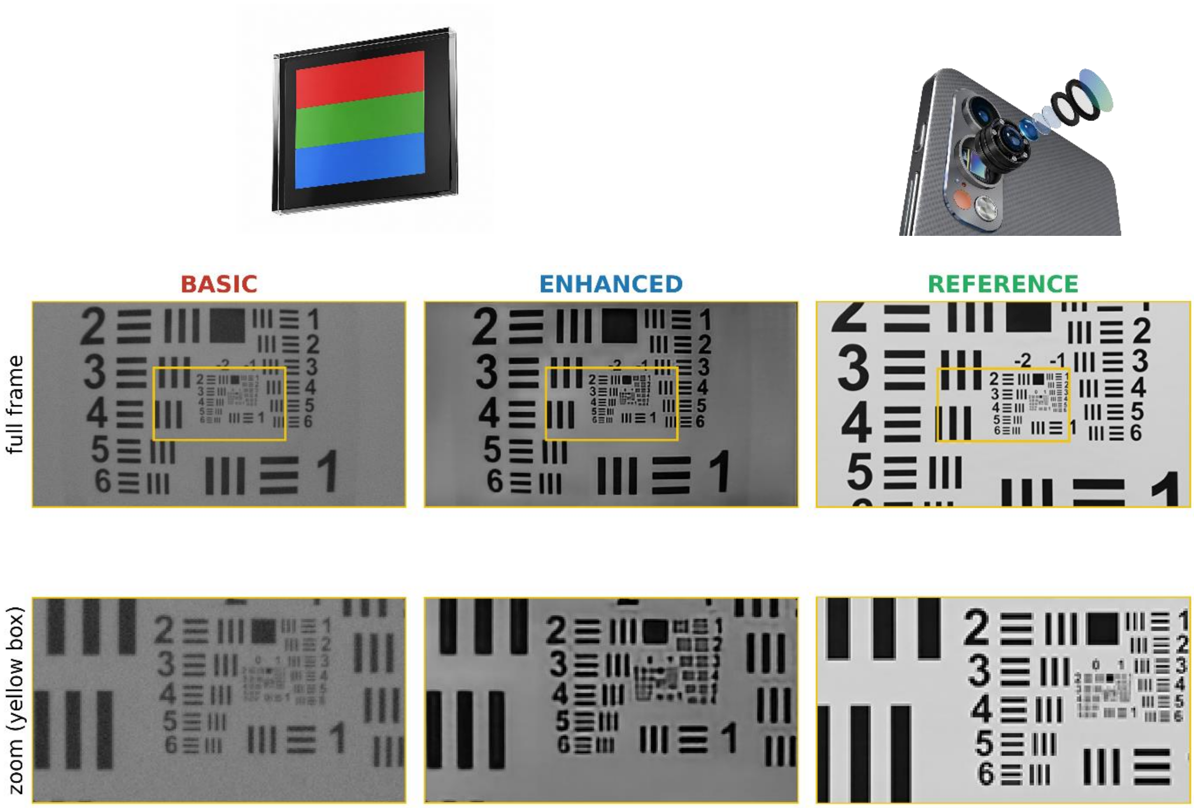


Figure 7: Comparison of single-layer rectangular metalens camera module versus a smartphone camera module with a conventional multi-lens stack. Left: Basic processing including brightening and white balance. Center: Basic processing with additional AI image enhancement. Right: smartphone camera image which leverages on a comprehensive processing stack including sharpening and HDR (see Fig. 9).

The benchmark target is a USAF-1951-style resolution chart, photographed handheld with both cameras. Because a good enhancement deliberately departs from its input, full-reference fidelity metrics are not applicable; all quality measures are therefore reference-free, computed per frame and averaged over 5 basic-processed and 5 AI-enhanced metalens frames. Sharpness is measured with an ISO-12233-style slanted-edge analysis on the large solid square of the chart, which yields the modulation transfer function (MTF). The two cameras sample the same chart very differently — the square spans roughly 57 pixels in the metalens frames but about 494 pixels in the native smartphone frame — so all spatial frequencies are reported in cycles per square width, a chart-referenced unit that is directly comparable across both systems. The smartphone image is down-sampled only for display in Figure 7; every measurement is made on its native full-resolution file.

Figure 7 shows the visual result. The basic-processed metalens frame is soft, grainy and low in contrast. After AI enhancement, the same capture becomes cleaner and crisper, with near black-on-white bars and a smoother background. The smartphone image shows the same chart through a high-end computational pipeline and marks the external quality ceiling for this test.

Figure 8 summarises the quantitative measures. From basic to enhanced, every measure improves on every frame: the dynamic range roughly doubles (78 to 160), RMS contrast rises about 1.6× (0.25 to 0.40), intensity entropy increases (5.1 to 6.5 bits), MTF50 rises about 1.7× (12 to 22 cycles per square width), the 10–90 % edge width halves (0.044 to 0.022 square widths) and the noise level falls about fourfold (3.1 to 0.7).

Two properties stand out as practically valuable. First, the model sharpens and denoises at the same time — a combination a linear filter cannot achieve, since linear sharpening necessarily amplifies noise. Second, it is robust to exposure: dark and normally exposed inputs converge to similarly high-quality outputs.

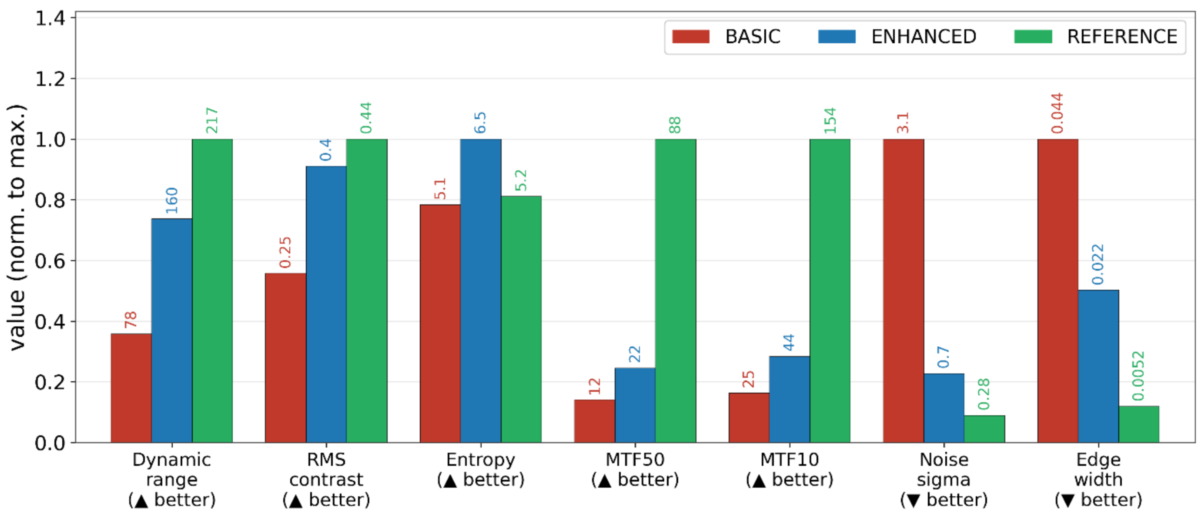


Figure 8: Seven image-quality measures for the basic-processed metalens frames, the AI-enhanced metalens frames and the smartphone reference, each measure normalised to its largest value, so the better result reaches 1.0; group means are printed above the bars and the arrows mark the favourable direction. MTF50 and MTF10 are given in cycles per square width and edge width in square widths; dynamic range, RMS contrast, entropy and noise are absolute. AI enhancement improves every measure relative to basic processing. The smartphone leads on most measures, with one exception: intensity entropy, where the enhanced metalens output slightly exceeds it.

These results place the single layer metalens camera on an honest footing. In raw resolving power the smartphone remains about 4× ahead — its MTF50 of roughly 88 cycles per square width is four times the enhanced value, and it continues to resolve detail well beyond the metalens camera's sampling limit.

On tone and contrast, however, the enhanced metalens output approaches the smartphone, and on intensity entropy it slightly exceeds it. What stands out is the asymmetry of the two imaging chains: the smartphone result is produced by a stack of refractive lens elements, a 50 MP sensor and a mature hardware and software ISP, while the metalens result is produced by a single nanostructured thin film on glass, a monochrome sensor and a software pipeline running on general-purpose hardware. That so simple an optical front-end, paired with learned restoration, delivers imaging within reach of a modern smartphone pipeline — with most of the processing acceleration described next still untapped — is the central performance message of this paper.

Processing Stack and Speed

The image quality shown above is produced in real time by MOTviewer, the MetaOptics software environment introduced earlier, in which classical processing nodes and AI models are combined into application-specific pipelines.

The runtime is compiled Cython with zero-copy buffer hand-off between nodes and releases the Python GIL during numerical work, so multiple cameras can run in parallel on a multi-core processor. AI filters can execute on GPU through the ONNX runtime with the CUDA execution provider; classical filters can use a CUDA-accelerated OpenCV plugin or are based on customized CUDA or SIMD optimized kernels. An example full model size is around 250MB, 17M Parameters, ~120 GFLOPs.

The model size is being optimized to reduce inference time for larger pixel resolutions. Currently, the combined system processes a one-megapixel frame in less than 50 milliseconds on a single off-the-shelf desktop GPU (NVIDIA RTX 3090), which is fast enough for real-time video.

Figure 9 places this stack in the wider landscape of edge image processing. A smartphone pipeline exploits nearly every acceleration layer available, from dedicated ISP silicon to neural processing units, while the MetaOptics pipeline today uses only a fraction of the available techniques — OpenMP, SIMD, OpenCV, CUDA and ONNX on CPU and GPU. Hardware ISP development and the use of neural or tensor processing units are the natural next steps, and each leaves substantial headroom for further gains in both speed and image quality.

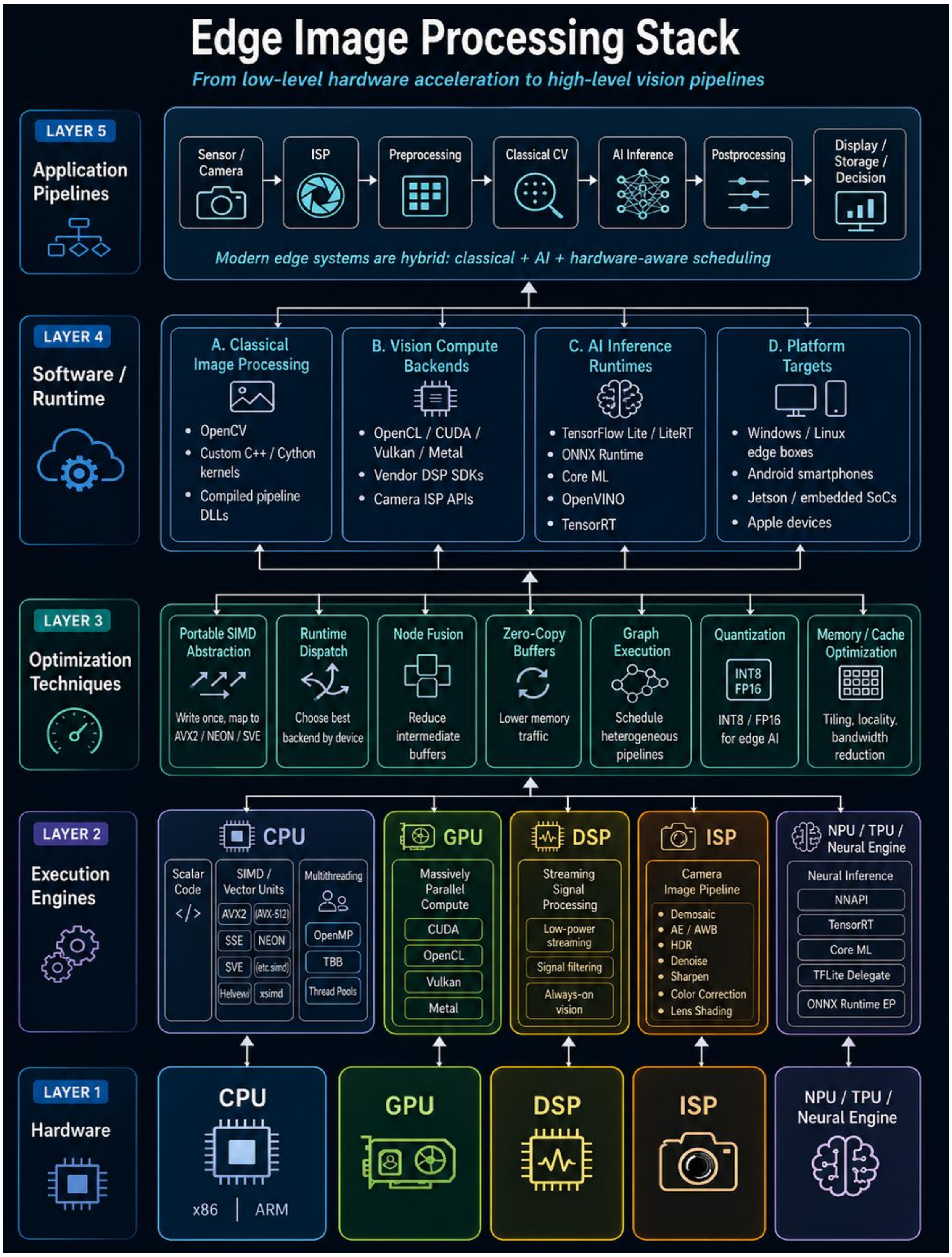


Figure 9: General edge image-processing stack. The diagram maps the path from hardware accelerators and execution engines to optimization techniques, software runtimes, and complete camera/image-processing pipelines. Metalens cameras typically only use a small fraction of possible image processing acceleration techniques compared to modern smartphone cameras. MetaOptics processing pipelines currently leverages on OpenMP, SIMD, OpenCV, CUDA and ONNX on CPU and GPU, while ISP development and use of Neural or Tensor Processing Units as well as a large range of optimization techniques leave huge room for further enhancements.

Applications

Image enhancement is a means to an end. The end is downstream applications that consume the enhanced image stream and turn it into something the customer can act on. MetaOptics ships three application heads as part of the platform today; all three benefit directly from the optical and software architecture described above.

Touchless Fingerprint Recognition

Conventional fingerprint readers require the user to press a finger against a glass platen — unhygienic, prone to wear, and unreliable on dirty or damaged fingers.

MetaOptics' touchless fingerprint module replaces the platen with a near-infrared metalens camera. The user holds a finger 3-5 cm in front of the sensor; the module captures the image, enhances ridge contrast, extracts feature points and matches against the stored template in real time. This working distance is a function of the lens-to-sensor distance and can be adjusted to suit specific needs.

The biometric module also inherits the RGB metalens free depth signal. The same fingertip is imaged simultaneously by the three metalenses arranged from different, known positions and the three resulting images carry slightly different perspectives of the same finger and ridge pattern. A genuine three-dimensional finger produces the curved disparity field expected from a curved surface; a flat photograph or a printed spoof held in front of the camera produces a flat disparity field that is geometrically incompatible with a real fingertip. The recogniser therefore validates not only the ridge pattern but also that the object presented to the camera is a real three-dimensional finger — built-in liveness detection at no additional sensor cost.

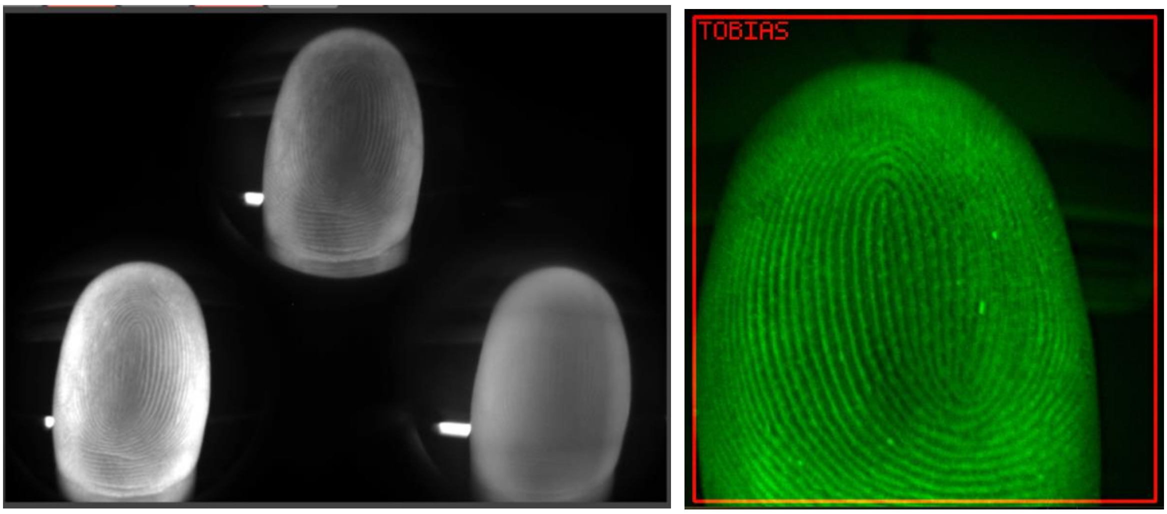


Figure 10: Touchless fingerprint capture and recognition through a MetaOptics metalens, with built-in three-dimensional liveness verification. Left: Live capture in MOTviewer — three simultaneous sub-aperture views of the same fingertip. The slight perspective differences between the three views are the depth cue that proves the object is a real three-dimensional finger, not a flat photograph. Right: After image enhancement and feature-point extraction, ready for matching against the stored template.

This "free" 3D detection capability is a big deal. In other biometric applications such as face detection for a computer, phone, border-control kiosks, door locks or automobiles, a Time-of-Flight sensor or dot projector is needed in addition to a metaoptic lens over the sensor. With three separate spatially separated regions (no color filters) capturing the same image, image processing can extract the parallax difference and model the 3D shape of the finger down to ridge details. This will save cost vs. the conventional approach.

This 3D detection capability is object distance dependent, however. The parallax is greatest for closer objects but may create issues with occlusion when reconstructing the image. As the object moves away from the sensor, parallax difference get smaller, but the occlusion issue gets reduced as well. For face identification, the spatial arrangement of the metalenses is expected to offer a significant cost advantage.

Gesture Recognition

Smart glasses, AR/VR headsets and many embedded vision products need a controller-free input modality. Hand gestures are the natural answer.

The gesture recognition node takes a color image from the metalens camera, detects the hand in each frame and classifies the gesture in real time. The current demo model recognises finger-counting gestures from zero through nine plus a null state — a discrete vocabulary sufficient for menu selection or numeric entry. Recognition accuracy is set as much by image quality as by the classifier; the upstream enhancement stages feed the recogniser a clean, well-lit, sharp image under variable ambient conditions.

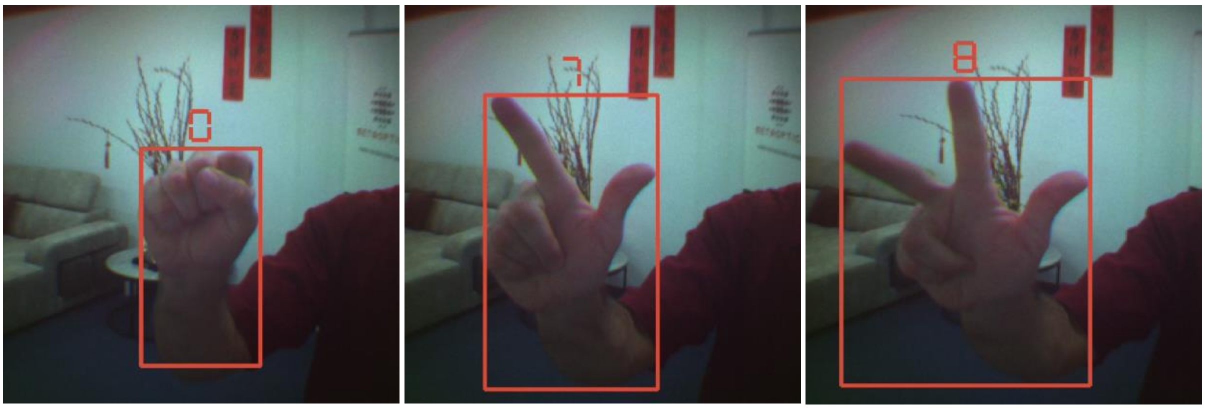


Figure 11: Real-time gesture recognition. Left: Closed fist: Digit zero. Center: Digit seven. Right: Digit eight.

From Prototype to Product

MetaOptics proprietary MOTviewer version allows a fast and seamless deployment of developed image processing pipelines. When combined with the camera module, developers have a robust development platform that can help transition to a mass-producible product.

The Pipeline Compiler analyses the active pipeline graph, transpiles it to optimised C and compiles a single Windows DLL with its dependencies. The library processes frames bit-identically to the live MOTviewer pipeline, with every filter setting — sigmas, kernel sizes, thresholds, calibrated channel positions — baked in at compile time. This capability is currently expanded to allow pipeline export to run on a variety of hardware solutions and operating systems.

An alternative complete SDK Export packages the same compiled pipeline with a system library that handles camera discovery, channel calibration and acquisition, plus a small C API and a ready-to-compile example. Our vision is that the customer’s host application simply links against the exported libraries, calls a handful of API functions and is in production.

Contact

MetaOptics Technologies Pte. Ltd.

81 Ayer Rajah Crescent, #01-45, Singapore 139967

sales@metaoptics.sg | www.metaoptics.sg

United States: MetaOptics Inc., 1 Ferry Building, Suite 201, San Francisco, CA 94111