

RESEARCH ARTICLE

Open Access



VEViD: Vision Enhancement via Virtual diffraction and coherent Detection

Bahram Jalali^{1*}  and Callen MacPhee¹

Abstract

The history of computing started with analog computers consisting of physical devices performing specialized functions such as predicting the position of astronomical bodies and the trajectory of cannon balls. In modern times, this idea has been extended, for example, to ultrafast nonlinear optics serving as a surrogate analog computer to probe the behavior of complex phenomena such as rogue waves. Here we discuss a new paradigm where physical phenomena coded as an algorithm perform computational imaging tasks. Specifically, diffraction followed by coherent detection becomes an image enhancement tool. Vision Enhancement via Virtual diffraction and coherent Detection (VEViD) reimagines a digital image as a spatially varying metaphoric “lightfield” and then subjects the field to the physical processes akin to diffraction and coherent detection. The term “Virtual” captures the deviation from the physical world. The light field is pixelated and the propagation imparts a phase with dependence on frequency which is different from the monotonically-increasing behavior of physical diffraction. Temporal frequencies exist in three bands corresponding to the RGB color channels of a digital image. The phase of the output, not the intensity, represents the output image. VEViD is a high-performance low-light-level and color enhancement tool that emerges from this paradigm. The algorithm is extremely fast, interpretable, and reduces to a compact and intuitively-appealing mathematical expression. We demonstrate image enhancement of 4k video at over 200 frames per second and show the utility of this physical algorithm in improving the accuracy of object detection in low-light conditions by neural networks. The application of VEViD to color enhancement is also demonstrated.

Keywords Physics-inspired computer vision, Natural algorithms, Low-light image enhancement, Color enhancement

1 Introduction

For two thousand years, the Antikythera mechanism lay quietly in the Mediterranean Sea, a timestamp of one of humanity’s first known attempts at artificial computing. It is theorized that the machine could calculate the positions of the sun and moon as a function of date and time [1]. Since then several other generations of computing machines were imagined and built, typically with the same continuous state space of this ancient device. Invented in 1206, Castle Clock was a hydro-powered

astronomical clock that was the first programmable analog computer [2]. Later, the industrial revolution saw the creation of analog machines that solve differential equations and calculate firing angles of artillery shells [3].

These devices perform a computational task by mapping it into a proxy mechanism that mimics the problem of interest. In this context, optics offers a unique platform for analog computing and realization of physical co-processors for the acceleration of scientific computing [3] such as emulation of Rogue Waves—a stochastically-driven nonlinear phenomenon [4, 5]. While analog computers utilize varying degrees of physical abstraction to model the actual system, there remains an underlying continuous space mapping between the states of the machine and the states of the system modeled.

*Correspondence:

Bahram Jalali
jalali@ucla.edu

¹ Electrical and Computer Engineering Department, UCLA, Los Angeles, USA

With the advent of much more predictable and governable digital devices, this mapping is violated, resulting in general-purpose computers that are tremendously successful in following any instructions coded in software. Given their theoretical and empirical performance bottleneck manifested in power dissipation and latency, the lure of faster, more efficient analog mappings for niche applications remains. Here we describe such a mapping, namely in the field of low-light image enhancement.

When captured in low-light conditions, digital images often incur undesirable visual qualities such as low-contrast, feature loss, and poor signal to noise ratio. The goal of low-light image enhancement is the abatement of these qualities for two purposes: increased visual quality for human perception and increased accuracy of machine learning algorithms. In the former, real-time processing can serve as a boon for convenient viewing, but in the latter, it serves as a requirement for emerging applications such as autonomous vehicles and security. Furthermore, video capture entails a fundamental tradeoff between light sensitivity, which is proportional to exposure time, and frame rate. This obviates the increase in exposure time as a meaningful solution to improving the image quality at low light levels because this would sacrifice the frame rate. In other cases, such as that of live-cell tracking in biology, image enhancement is crucial as low light conditions are necessary to avoid phototoxicity (cell death caused by exposure to light).

Considering the present computational landscape and constraints described above, we introduce a physics-inspired, real-time low-light image enhancement algorithm with a theoretical mapping to the physics occurring in natural systems in the analog domain. We show this algorithm has exceptional performance in terms of image quality and computational speed. In the Methods section, we explain the intuition behind the algorithm and deep insights on how it works.

1.1 Prior work on low-light level enhancement

There has been a great deal of progress in the task of low-light image enhancement in recent years, primarily due to the adoption of powerful machine learning approaches. We therefore split our brief discussion of prior work on low-light level enhancement into classical algorithms that are deterministic and machine learning approaches which are data-driven.

1.1.1 Classical algorithms

The field of low-light image enhancement has a very diverse solution set, with several classical algorithms of varying complexity and performance. While the field still lacks a unifying quantitative theory, Retinex theory has arisen as one of the mainstay concepts in classical

approaches. Stemming from concepts in human perception theory concerning decomposition of an image into an illumination and a reflectance constituent, Retinex based approaches account for a large portion of low-light image enhancement techniques [6–8]. LIME [9] is one such algorithm that utilizes optimized Retinex theory for illumination map generation for high-quality enhancement. Among classical algorithms, it shows very high performance over a large range of lighting conditions [6]. Similarly, histogram equalization [10] methods are a widely used alternative that create an expanded, more uniform histogram for contrast enhancement and increased dynamic range, yet these methods often suffer from color distortion and other artifacts [11]. To diminish these qualities, local and adaptive histogram equalization techniques have been proposed such as CLAHE [11]. Several other classes of traditional algorithms include frequency-based, defogging, and image fusion methods that are used in High Dynamic Range (HDR) techniques.

1.1.2 Deep learning approaches

The proliferation of deep learning algorithms in the last decade has touched many different fields, and image enhancement is no exception. The preponderance of novel algorithms within the field have been data-driven. On the side of supervised learning, one of the first deep learning based approaches, LLNet [12], gave rise to many other autoencoder based designs. Other networks, like MBLLEN [13], EEMEFN [14], TBEFN [15], all make use of similar ground truth datasets for training. In comparison, networks such as Retinex-Net are built upon the theoretical underpinnings of Retinex's human perception theory and therefore are more interpretable.

All these approaches demonstrate high-performance in target lighting conditions, but typically they have difficulty generalizing to greater domains not covered within training data. Other neural network approaches utilize unsupervised learning in the form of generative models such as EnlightenGAN [16]. Lastly, zero-shot techniques that do not require labeled data, such as Zero-DCE [17], have shown good image quality and fast inference speeds. In Zero-DCE, a group of equalizing *s*-curves are generated at inference time. These curves are learned through a training process that utilizes a set of custom no-reference loss functions that compute several enhancement characteristics such as exposure error and spatial consistency. While this approach needs no ground truth (labeled training data) it still requires training time and diverse image data. Owing to its small network size, however, the network has fast inference time, making it a candidate for real-time image enhancement at certain resolutions.

While these blackbox machine learning models have been revolutionary, they ultimately are restricted by the

accuracy of their loss functions in the absence of labeled reference data. As low-light image enhancement still lacks a rigorous quantitative loss function that accurately reflects human perception, these approaches don't perform well when such heuristic metrics fail to correctly define the enhancement as it would be perceived by the user. In other words, the algorithms may produce images that satisfy the minimum loss function requirement but are not perceived as good images by a human viewer.

In this paper we introduce a new low-light level enhancement computer vision algorithms that is derived from the processes of propagation and detection of light. The algorithm emulates the propagation of light through a physical medium with engineered diffractive properties followed by coherent detection. Unlike traditional algorithms that are a sequence of hand-crafted empirical rules or learning based method that are trained and lack interpretability, our physics-inspired approach leverages

a law of nature as a blueprint for crafting an algorithm. Such algorithm can, in principle, be implemented in an analog physical device for fast and efficient computation.

2 Vision Enhancement via Virtual diffraction and coherent Detection (VEViD)

2.1 Physics framework

Ubiquitous in nature as well as in optical imaging systems, electromagnetic diffraction is a process in which light acquires a frequency-dependent phase upon propagation. The phase increases with spatial frequency and in the paraxial approximation, it is a quadratic function of frequency. While the human eye and common image sensors respond to the power in the light, instruments can work with both the intensity and phase of light, with the latter being measured through coherent detection.

Vision Enhancement via Virtual diffraction and coherent Detection (VEViD) introduced here reimagines a

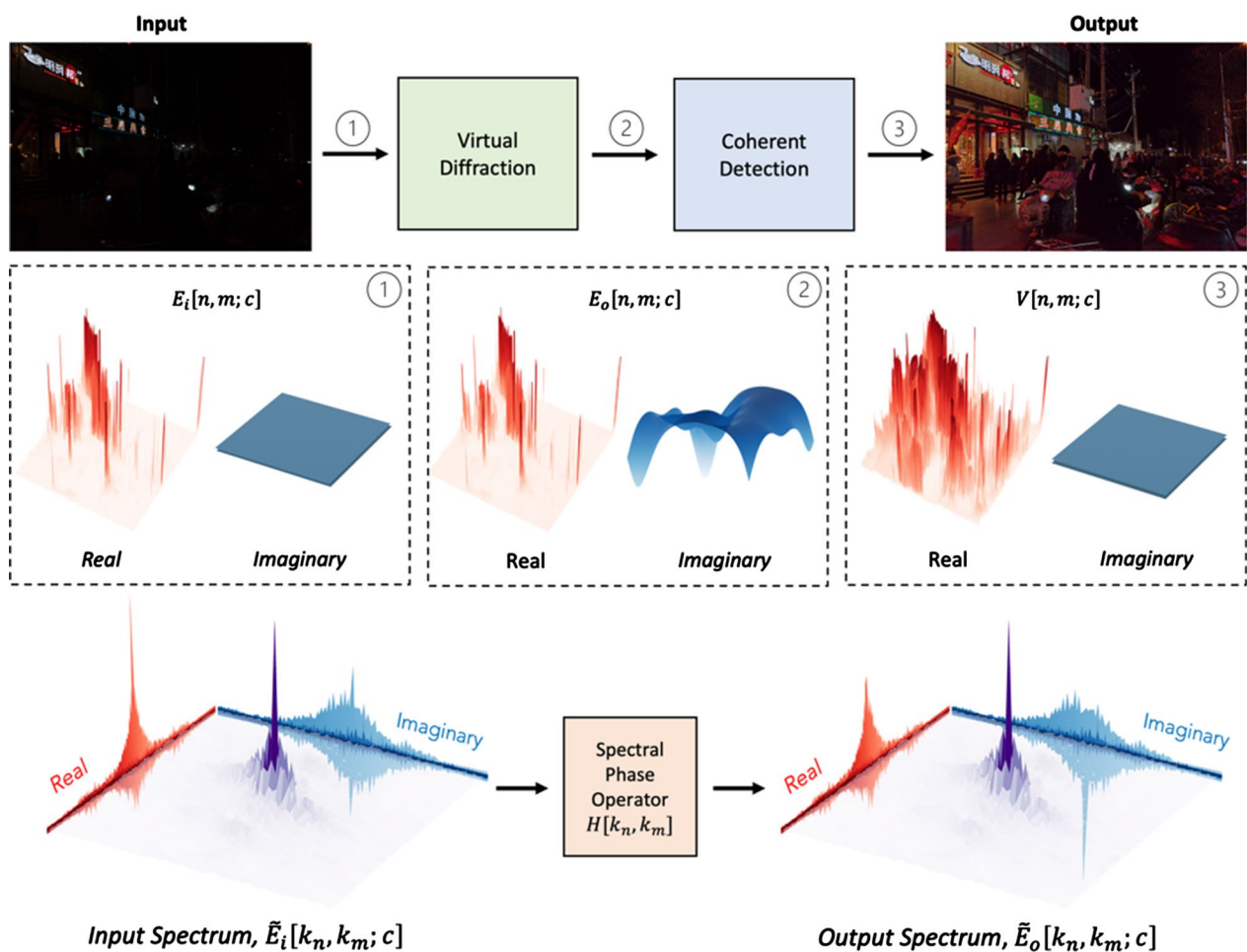


Fig. 1 Physical interpretation of the VEViD algorithm showing its impact in spatial domain (top row) and in spectral domain (bottom row). In spatial domain, the real part of the image is nearly unchanged whereas an imaginary part is created after diffraction. This observation supports the mathematical approximation in the latter part of the paper

digital image as a spatially varying metaphoric light field. It then subjects the field to the physical processes akin to diffraction and coherent detection. The term “Virtual” captures the deviation from the classical diffraction. The virtual world deviates from the physical world in three aspects. The light field is pixelated, and the propagation imparts a phase with an arbitrary dependence on frequency which can be different from the monotonically-increasing behavior of physical paraxial diffraction. Temporal frequency is restricted to three color bands.

To describe this process, we start with the general solution to the homogeneous electromagnetic wave equation in rectangular coordinate (x, y, z)

$$E(x, y, z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{E}_i(k_x, k_y, 0) e^{+jk_z z} e^{i(k_x x + k_y y)} dk_x dk_y \quad (1)$$

where $\tilde{E}_i(k_x, k_y, 0)$ is the spatial spectrum of the input field $E_i(x, y, 0)$. Then the Fourier content of the signal after a distance z gains a phase term which can be represented by a spectral phase, $\phi(k_x, k_y)$,

$$\tilde{E}_0(k_x, k_y, z) = \tilde{E}_i(k_x, k_y, 0) e^{-i\phi(k_x, k_y)} \quad (2)$$

The phase represents the total accumulated over the propagation length. We may rewrite the forward propagated signal subjected to diffractive phase as,

$$E_o(x, y, z) = IFT \left\{ \tilde{E}_i(k_x, k_y, 0) e^{-i\phi(k_x, k_y)} \right\} \quad (3)$$

where IFT refers to the inverse Fourier transform. $E(x, y, z)$ now contains frequency-dependent phase profile that is entirely described by our arbitrary phase $\phi(k_x, k_y)$. The propagation converts a real-valued input $E_i(x, y, 0)$ to a complex function $E_o(x, y, z)$. As described below, we are interested in the phase of this complex function.

As we are concerned with digital images, we now move from a continuous valued $E(x, y)$ in spatial domain to discrete, meaning pixelated, waveform $E[n, m]$. Similarly in the frequency domain from continuous (k_x, k_y) to discrete momentum $[k_n, k_m]$.

Of primary interest to us is the “lightfield”, which we define as the distribution of “field” strength across the two-dimensional landscape of the input signal with the pixel brightness mapped into the metaphoric field strength. The equivalent temporal frequency of the lightfield has three bands corresponding to the three fundamental color channels (RGB). To arrive at the field for color images, we transform our input RGB image into the hsv color space. We will refer to this quantity as $E[n, m; c]$ where c is the index for the color channel To preserve the color integrity, the diffractive transformation operates

only on the “v” channel of the image when performing low-light enhancement.

2.2 Mathematical framework

For results that follow, spectral phase filter has a low pass characteristic. A wide range of low pass spectral phase functions can be used. While it may not be the optimum function, for simplicity a Gaussian function with zero mean and variance T for the frequency dependent phase is considered here,

$$\phi[k_n, k_m] = S \cdot \exp \left[-\frac{k_n^2 + k_m^2}{T} \right] = S \cdot \hat{\phi} \quad (4)$$

Resulting in a spectral phase operator,

$$H[k_n, k_m] = e^{-i\phi[k_n, k_m]} \quad (5)$$

where S is a model parameter that maps into propagation loss (or gain). In the physical wave propagation, the spectral phase induced by diffraction depends on the propagation length. In VEVid, the length is reflected in the phase scale parameter, S . The value of S and hence the propagation length are constrained by the requirement that the propagation induced phase must be small. We refer to this regime as “nominal near field” as explained in the Methods section.

Following the application of the spectral phase and inverse Fourier transform, coherent detection produces the real and imaginary components of the field from which the phase is obtained. The combined processes of diffraction with the low pass spectral phase and coherent detection produces the output of VEVid, $V[n, m]$,

$$V[n, m; c] = \text{angle} \langle IFT \{ e^{-i\phi[k_n, k_m]} \cdot FT \{ E[n, m; c] \} \} \rangle \quad (6)$$

where FT denotes the Fourier transform, and the angle operator calculates the phase of the complex-valued function of its argument. Previously, other types of spectral phase operations have been exploited in creating edge detection algorithms [18, 19].

2.3 Impact of VEVid in spatial and frequency domains

The effect on the spatial domain representation is shown in Fig. 1 (top row). The input image is a real valued function. After virtual diffraction, real component is nearly unchanged however the image acquires a significant imaginary component. After phase detection, the image is once again a real valued function but is significantly different from the input.

The effect on the spatial frequency domain is shown in Fig. 1 (bottom row). The imaginary portion of the spectrum adopts a central low frequency spike, while the real

portion undergoes corresponding attenuation in its low frequency component due to energy conservation.

We point out again that the Gaussian spectral phase function was chosen for its mathematical convenience and further performance gains may be possible by exploring other phase profiles. Later in the computational acceleration section, we show that the phase can be approximated by a constant when combined with other approximations. Also, while the low pass spectral phase does not occur in natural diffraction (where phase increases with spatial frequency), it can be synthesized by spatial light modulators (SLM) or metamaterial combined with diffractive optics.

2.4 The VEVid algorithm

The VEVid algorithm is formally defined in Fig. 2. The input image is first converted from RGB to *hsv* color space (not shown). For low light level enhancement the hue channel and saturation channels are not transformed because we seek to retain the original color mapping of the input image. For color enhancement, the *v* and *h* channels are left unchanged and the *s* channel is transformed with VEVid.

A small constant bias term, *b*, is added to the field for the purposes of numerical stabilization and noise reduction. This step is not necessary but it improves the results. The real-valued input image is then transformed into the Fourier domain by FFT and subsequently multiplied elementwise by the complex exponential with an argument which defines the frequency dependent phase. Inverse Fourier transform (IFFT) returns a complex signal in the spatial domain. Mathematically, the inverse tangent operation in phase detection behaves like an activation function. Before computation of phase, the signal is multiplied by a parameter *G* called phase activation gain. The output phase is then normalized to match the image formatting convention [0 to 255]. This output is

then injected back into the original image as the new *v* channel (for low light enhancement) or the *s* channel (for color enhancement).

The results shown in Fig. 3 demonstrate the quality and generalization of VEVid to several application domains and illumination conditions. We note the ability of VEVid to produced enhanced images with natural colors.

Along with low-light image enhancement, the VEVid transformation is also capable of performing color enhancement for realistic tone matching when applied to the saturation channel of the input image. The process is identical to that of the low-light enhancement procedure described previously with the exception the transform is applied to the saturation channel. The results are shown in Fig. 4.

The benefits of having a physical algorithm include low computational burden owing to its simplicity, generalizability to wide range of domains, and the potential for implementation in the analog (physical) domain using diffractive optics.

2.5 Application to object detection via deep neural networks

With the advent of deep learning, computer vision approaches are having spectacular success in applications such as autonomy, manufacturing, and security and defense. On the other hand, these approaches are often unpredictable in complex real-world environments that involve heterogenous data and outliers not represented within the training set. In this section, we demonstrate how pre-processing with VEVid maps image data into a form which improves the accuracy of object detection using off the shelf neural network algorithms without having to retrain them on low light conditions.

The amount of time, memory, and energy required to train a deep neural network, store and recall the millions or billions of model parameters is expensive and is

Algorithm 1: VEVid: Vision Enhancement via Virtual diffraction and coherent Detection

Data: $E_i[n, m; c]$, *input image*

Result: $V[n, m; c]$, *output image (VEVid transform)*

$$\phi[k_n, k_m] \leftarrow S * \hat{\phi}[k_n, k_m]$$

$$H[k_n, k_m] \leftarrow \exp(-i * \phi[k_n, k_m])$$

$$E_i[n, m; c] \leftarrow E_i[n, m; c] + b$$

$$\tilde{E}_i[k_n, k_m; c] \leftarrow FFT\{E_i[n, m; c]\}$$

$$\tilde{E}_o[k_n, k_m; c] \leftarrow \tilde{E}_i[k_n, k_m; c] * H[k_n, k_m]$$

$$E_o[n, m; c] \leftarrow IFFT\{\tilde{E}_o[k_n, k_m; c]\}$$

$$V[n, m; c] \leftarrow \tan^{-1}\left(G * \frac{Im\{E_o[n, m; c]\}}{Re\{E_o[n, m; c]\}}\right)$$

$$V[n, m; c] \leftarrow N\{V[n, m; c]\}$$

Fig. 2 The VEVid algorithm

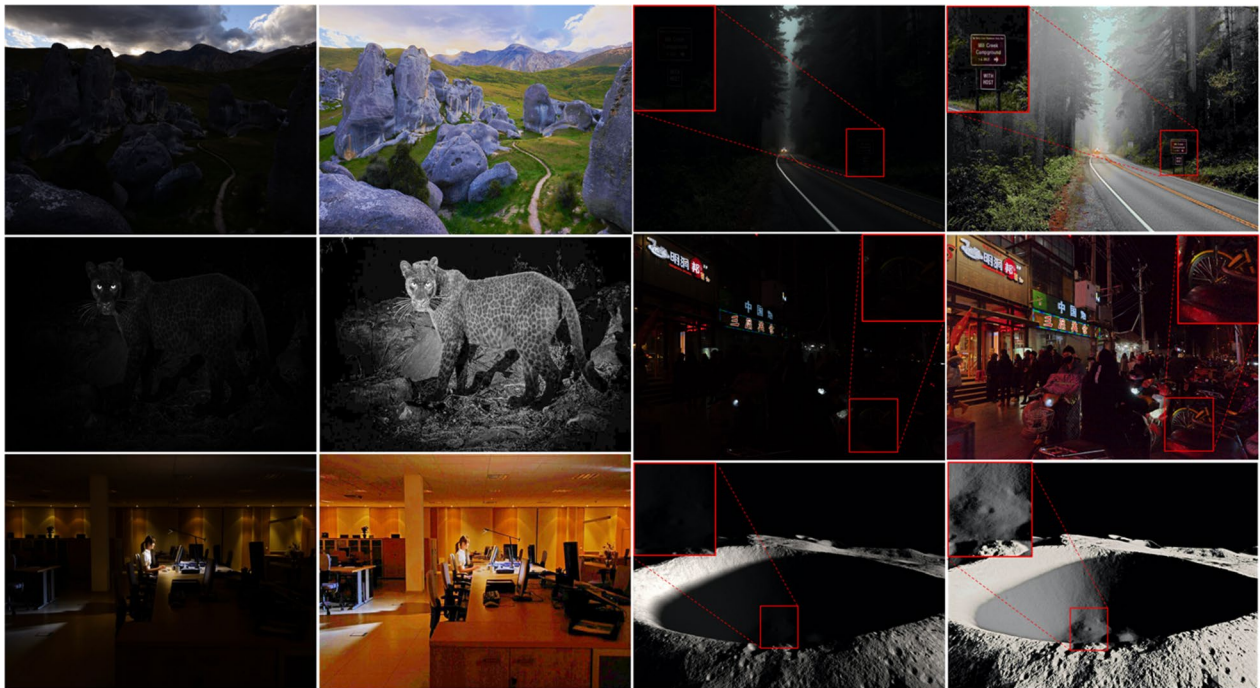


Fig. 3 Demonstrating applications of VEVid to various types of images. (Left image pairs) Examples of VEVid’s enhancement of images captured in low-light conditions. (Right image pairs) VEVid’s ability to enhance previously invisible details in low-light images. VEVid parameter values: $S=0.2$, $b=0.16$, $G=1.4$. The same set of values works over a wide range of images and illuminations

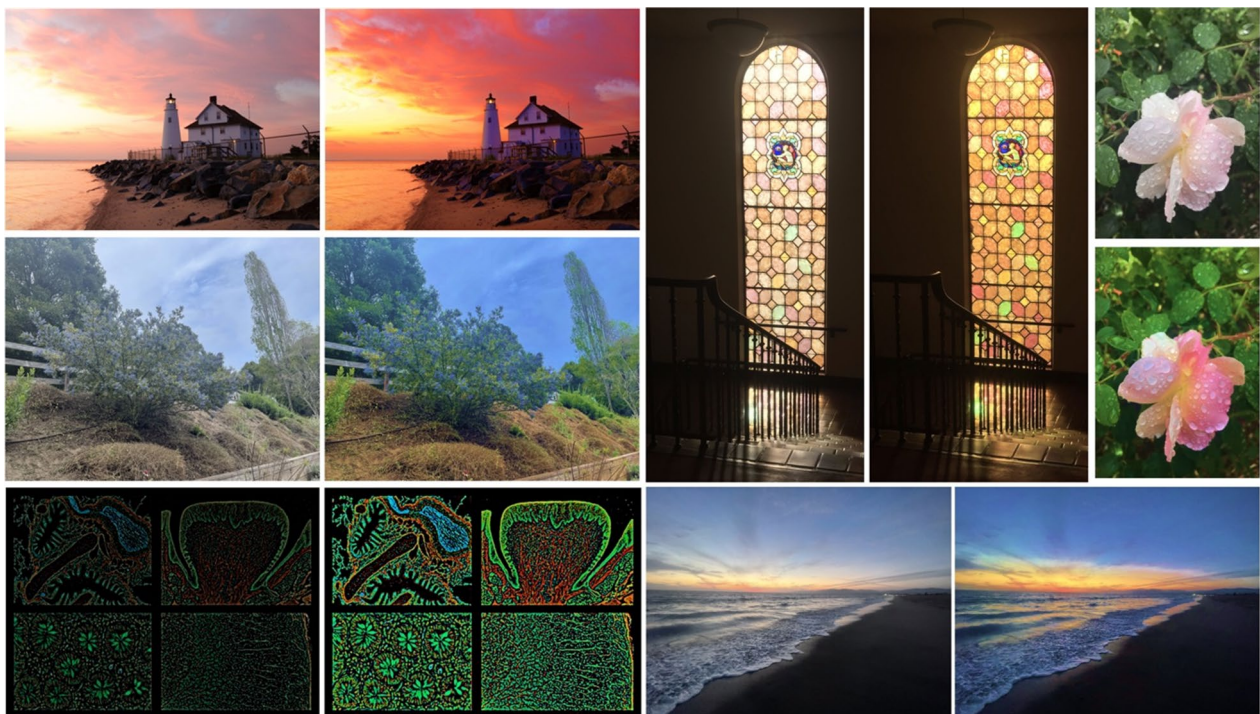


Fig. 4 Color enhancement using the VEVid algorithm. Here VEVid operates on the saturation channel of HSV input image. VEVid parameter values: $S=0.4$, $b=0.20$, $G=1.5$

outpacing the growth in semiconductor performance as described by the Moore's Law [20]. We take for example a powerful object detection neural network, YOLO [21], which has tens of million parameters. While the of ability of such networks to learn complex patterns in the data is vast, the efficacy comes down to the millions of free parameters in the model and the size and richness of the dataset available to fit those parameters. In addition, retraining of a network such as this for domain-specific applications such as low light conditions is burdensome as it requires large new datasets that must be acquired plus the additional training cost. For an application such as pedestrian detection, it is very important that YOLO generalizes well to low-light conditions, especially in cases such as autonomous driving. Another important application is the security camera. Unless the network is presented with labeled training images captured under low light levels, there is no guarantee that it will function correctly. Preprocessing the image with an algorithm such as VEVid represents a way to easily increase the generalization of these networks to low light level

conditions. We see in Figs. 5 and 6 the increased performance of the YOLO pretrained model when the image is first processed by VEVid. In each image, VEVid identifies previously obscure features within the images. In addition, there are several objects missed by the pretrained YOLO network that, after preprocessing with VEVid, are now detected.

Such increased detection accuracy is of vital importance to applications such as autonomous vehicles and security camera systems. In these cases, performance must generalize to night-time environments. In response to this challenge, we have shown that VEVid increases the performance of state-of-the-art neural network inference in such environments. In the next section, we demonstrate VEVid's exceptional computational efficiency—a key attribute for real world applications where low latency is crucial.

2.6 Video rate object recognition in low-light conditions

Video applications provide a rich environment for low-light image enhancement due to the tradeoff between

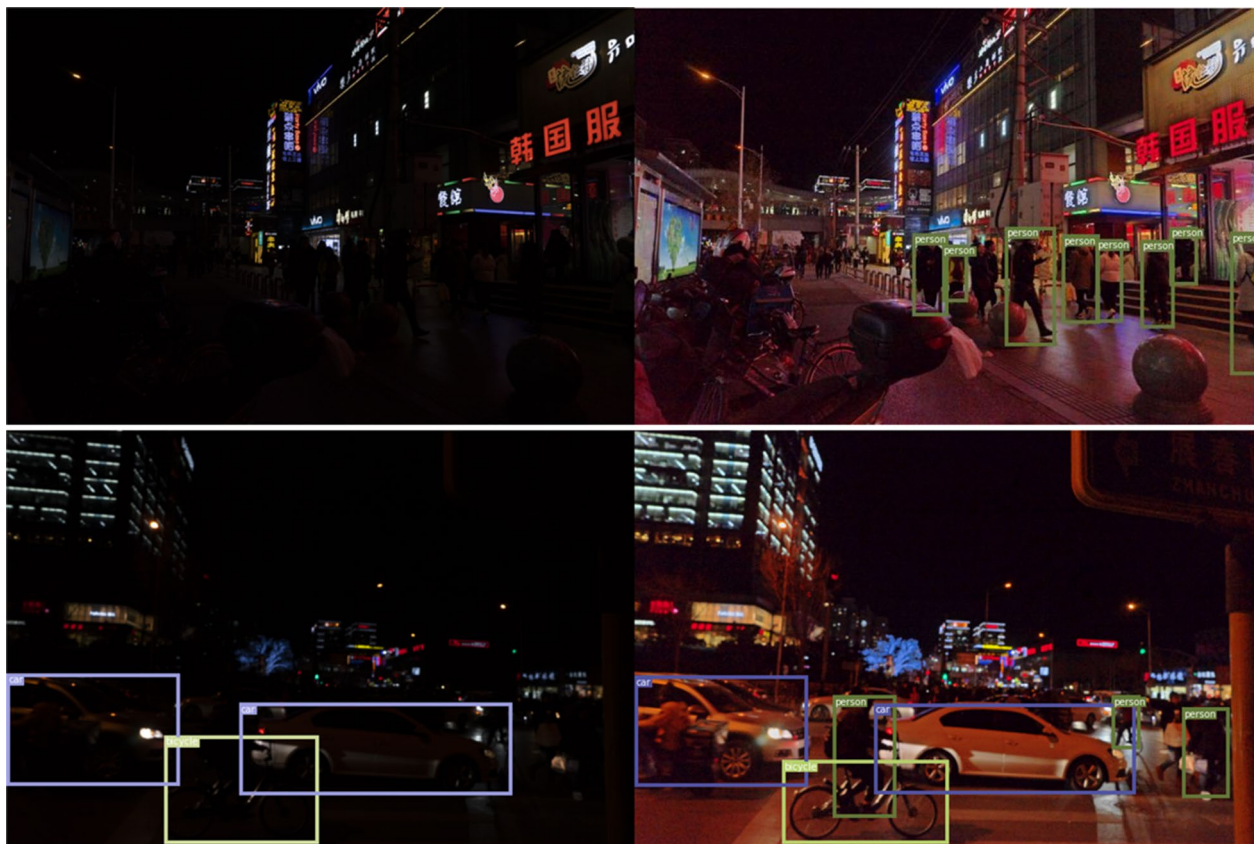


Fig. 5 Impact of VEVid preprocessing on object detection. Examples of object detection using the leading AI neural network (YOLO) on low-light images without (left) and with (right) preprocessing with VEVid. Here, objects recognized by the neural network are shown by bounding boxes. The object detection algorithm is a pretrained YOLOv3 network. VEVid parameter values: $S=0.2$, $b=0.16$, $G=1.4$



Fig. 6 YOLO performing object tracking on a security camera video on original (left) and VEVid preprocessed images (right). Image from an infrared camera (center) shows that hidden details that are revealed by VEVid match the real scene. $S=0.4$, $b=0.16$, $G=1.4$

image quality and frame rate. Capturing a video at a high frame rate and without blurring requires short integration time (fast shutter). Such low integration time in turn leads to poor image quality at low light levels because fewer photons are collected. Image enhancement can reduce this constraint, with the caveat that in real-time applications, the enhancement procedure itself must be fast enough as to not slow the frame rate. As we will see below, VEVid performs real-time low-light image enhancement at much higher frame rate than a

state-of-the-art neural network technique while producing comparable or better image quality.

Figure 7 left panel shows the runtime vs. the image frame size for VEVid as performed on an NVIDIA GeForce GTX TITAN X Graphic Processing Unit (GPU). Such asynchronous runtimes are measured using specialized timing functions within the PyTorch library (see the Methods section). The VEVid algorithm operates in real-time at a frame rate of 24 FPS past 4K video (8,294,440 Mega pixels). Shown for comparison is the performance

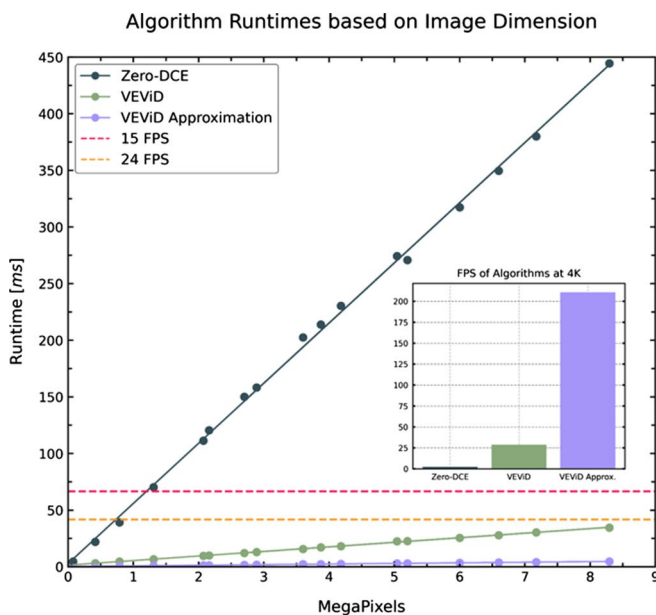


Fig. 7 Left: Comparison of runtime for the VEVid and the Zero-DCE algorithms for a range of image sizes. VEVid is capable of processing 4K resolutions images (8,294,440 Mega pixels) at 24 FPS and scales much better with frame size. The VEVid approximation developed for computational acceleration is even faster with capability of processing 4K resolution frames at more than 200 FPS. The approximation is a closed-form equivalent-model of the numerical VEVid algorithm that avoids the Fourier transform operations. **Right:** Comparison of Zero-DCE (middle) and VEVid (bottom) for two diverse input images (top). VEVid Parameters for the pedestrian and lamp images: $S=0.2$, $b=0.16$, $G=1.4$ and $S=0.2$, $b=0.32$, $G=1.4$, respectively

of Zero-DCE, a state-of-the-art deep learning algorithm with the shortest inference times according to a recent survey [22]. The survey also shows that Zero-DCE compares favorably with other state-of-the-art algorithms in terms of image quality. VEVid scales better with frame size than Zero-DCE with the advantage becoming dramatic for 4K frames. The right panel in Fig. 7 compares image quality between Zero-DCE and VEVid. The top figure is the input, the middle shows the Zero-DCE and the bottom is VEVid. Additional comparisons with Zero-DCE are provided in Fig. 8. Both algorithms perform well. Zero-DCE performs better in the cloudy regions of the images where the input image has high brightness whereas VEVid provides brighter and more intense images.

The fast runtime implies that VEVid can be inserted into the camera ISP as a preprocessing step for video applications without sacrificing frame rate. These results show the potential to augment real-time neural network based classification algorithms such as YOLO so their inference performance generalizes to low illumination conditions with no need for additional training data.

Whereas the full VEVid algorithm enables high-quality enhancement with state-of-the-art computational performance, in the next section we develop a framework for even faster performance through a mathematical approximation. The mathematically accelerated “VEVid-Lite” enables blazing-fast speed with limited penalty in image quality. The resulting equivalent model of VEVid is derived below followed by demonstration of its performance.

3 Computational acceleration

Low latency is a crucial metric for realtime applications including video analytics and broadcast. We are motivated to investigate whether VEVid can be further accelerated through mathematical approximations that reduce the computation time without appreciable sacrifice in image quality. In essence, we are seeking a compact closed-form equivalent model for VEVid. In doing so, we draw inspiration from the field of semiconductor device modeling where complex device physics is approximated as simple, albeit empirical, closed form equations



Fig. 8 Additional comparisons with Zero-DCE showing how the other images used in the paper compare. Both algorithms perform well. Zero-DCE performs better in the cloudy regions of the images where the input image has high brightness whereas VEVid provides brighter and more intense images

enabling fast simulations of complex circuits consisting of a massive number of those devices [23].

As shown in the left panel of Fig. 7, this approach, which is described below, leads to significant acceleration of the algorithm enabling processing of 4K frames at over 200 fps. Furthermore, as shown in Fig. 9, the quality of the output remains high and works well as a preprocessing step for enhancing the accuracy of object detection in low light conditions. Below, we provide the derivation of this simplified equivalent model for the VEViD.

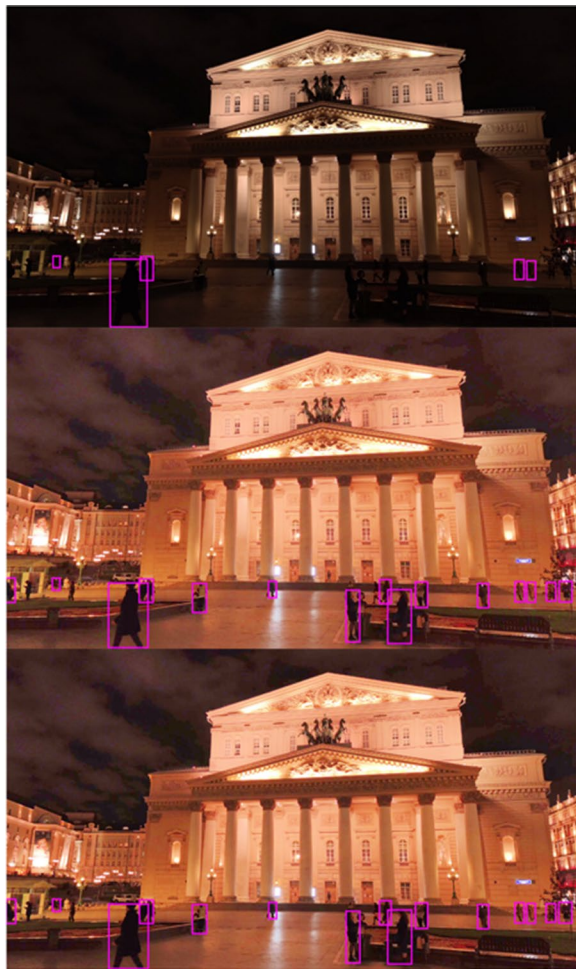


Fig. 9 Impact of preprocessing an image with VEViD on object detection by a neural network (YOLO). When applied to the original image, YOLO identifies 5 objects. After preprocessing by VEViD, the same YOLO algorithm detects 15 objects without having to be retrained on low-light images. Middle image is preprocessed by the full VEViD whereas the bottom image is preprocessed with its simplified approximation (VEViD-Lite). The approximation has very similar visual quality but with much lower latency (Fig. 7). Parameter values for the full VEViD (middle image): $S = 0.4$, $b = 0.5$, $G = 0.6$, and for the VEViD approximation model (bottom image): $b = 0.5$, $G = 0.6$

The most time intensive operations in VEViD are the forward and inverse Fourier transforms. If we can find an equivalent formulation that takes place entirely in the spatial domain, this would avoid the Fourier transform and significantly improve the runtime of the algorithm. This will significantly reduce the latency enabling real-time enhancement of high resolution and high frame rate videos.

The mathematical simplification of the VEViD algorithm is enabled by three approximations. First, we assume that the real part of the image is not appreciably affected by diffractive propagation, as evident in Fig. 1. Second, by assuming the phase angle induced by virtual diffraction to be small, we remove the inherent nonlinearity of the complex exponential of the phase function. Third, by assuming the spectral phase to be a constant, the Fourier transform operation is avoided. We note that the price paid for the mathematical simplification is that we now depart from direct physical foundation of the algorithm. Despite the approximations and the resulting computational acceleration, the deeply simplified algorithm delivers excellent low light level enhancement, compared to the full version of the VEViD, as seen in Fig. 9.

We start from the output of the VEViD algorithm (Fig. 2),

$$V[n, m; c] = \tan^{-1} \left(G * \frac{\text{Im}\{E_o[n, m; c]\}}{\text{Re}\{E_o[n, m; c]\}} \right) \quad (7)$$

The original image is a real-valued quantify with no imaginary component. The spectral phase induced by diffraction produces an imaginary component but the change in the real component is negligible, as shown in Fig. 1. We therefore approximate the real component with the input,

$$\text{Re}\{E_o[n, m; c]\} \approx E_i[n, m; c] \quad (8)$$

resulting in the first simplification of the VEViD output,

$$V[n, m; c] = \tan^{-1} \left(G * \frac{\text{Im}\{E_o[n, m; c]\}}{E_i[n, m; c]} \right) \quad (9)$$

As a side note, the property of VEViD that equalizes the illumination can be understood by interpreting Eq. 9 as follows. The division by the input image, $E_i[n, m; c]$, in the argument of the arctan function emphasize the low intensity regions of the image producing low-light enhancement. Subsequently, the arctan operation compresses the output preventing an undesirable dynamic range expansion and suppressing the noise. Together these operations redistribute the energy while managing the dynamic range and noise.

We now focus on the imaginary component appearing in the numerator. Further simplification can be

made by linearizing the complex exponential operation encountered in the spectral phase (Eq. 6). This is done by restricting the phase to be small (nominal near field),

$$\begin{aligned} \exp(i * \phi[k_n, k_m]) &= \cos(\phi[k_n, k_m]) - i * \sin(\phi[k_n, k_m]) \\ &\approx 1 - i\phi[k_n, k_m] \end{aligned} \quad (10)$$

This leads to the following expression for the imaginary component:

$$\begin{aligned} \text{Im}\{E_o[n, m; c]\} &= \text{Im}\{IFT[FT(E_i[n, m; c]) * \exp(-i * \phi[k_n, k_m])]\} \\ &\approx \text{Im}\{IFT[FT(E_i[n, m; c]) * 1 - i * \phi[k_n, k_m]]\} \\ &= -IFT[FT(E_i[n, m; c]) * \phi[k_n, k_m]] \end{aligned} \quad (11)$$

As previously mentioned, the main effect of the spectral phase induced by diffraction is to produce an imaginary component. The real part of the output is a bright-field image with a large initial value, whereas the imaginary part is a dark-field image which is zero before diffraction. Any numerical noise will affect the imaginary part far more than the real part. To avoid this effect we regularize the imaginary component with a constant, b ,

$$E_i[n, m; c] \rightarrow E_i[n, m; c] + b\# \quad (12)$$

The final step in obtaining the simplified equivalent model for VEVid is to eliminate the Fourier transforms. This occurs in the limit where the phase variance, T , approaches infinity,

$$\lim_{T \rightarrow \infty} \phi[k_n, k_m] = S * \exp\left[-\frac{k_n^2 + k_m^2}{T}\right] \approx S \quad (13)$$

applying this to the imaginary component (numerator) of the diffracted image leads to the elimination of the Fourier and inverse Fourier transform operations,

$$\begin{aligned} \lim_{T \rightarrow \infty} \text{Im}\{E_o[n, m; c]\} &= \lim_{T \rightarrow \infty} (-IFT[FT(E_i[n, m; c] + b) * \phi]) \\ &= -S * (E_i[n, m; c] + b)\# \end{aligned} \quad (14)$$

Combining these steps leads to a simple closed form formulation of the VEVid algorithm,

$$\begin{aligned} \lim_{T \rightarrow \infty} V[n, m; c] &= \lim_{T \rightarrow \infty} \tan^{-1}\left(G * S \cdot \frac{\text{Im}\{E_o[n, m; c]\}}{\text{Re}\{E_o[n, m; c]\}}\right) \\ &= \tan^{-1}\left(G * \frac{-E_i[n, m; c] + b}{E_i[n, m; c]}\right)\# \end{aligned} \quad (15)$$

where the G parameter is redefined in the final step as the product of $G * S$ (new G absorbs S).

Avoiding the Fourier transform operations, Eq. 15 is a computationally-accelerated reformulation of VEVid. Figure 9 shows that this computationally-accelerated

VEVid approximation compares very favorably with the full numerical version of the algorithm in both the visual quality and as a preprocessing step for improving the accuracy of object detection by YOLO under low light conditions.

4 Conclusion

Physical diffraction and coherent detection can be used as blueprints for the transformation of digital images and videos leading to a new and surprisingly powerful algorithm for low-light and color enhancement. Unlike traditional algorithms that are mostly hand-crafted empirical rules, the VEVid algorithm presented here emulates physical processes and adapts them to the low-light level enhancement of digital images. In contrast to deep learning-based approaches, this technique is unique in having its roots in deterministic physics. The algorithms are therefore interpretable and do not require labeled data for training. Although the mapping to physical processes is not precise, in the future it may be possible to implement a physical device that executes the algorithm in the analog domain.

We demonstrated low-light enhancement with image quality comparable to the state-of-the-art neural networks but with much lower latency. While the full VEVid algorithm enables high-quality enhancement with high computational speed, we also developed a framework for even faster speed through a mathematical approximation. This enables low-light enhancement on 4k video at 200 frames per second. There are only two model parameters, G and b .

We also like to point out some of the limitations of the present version of VEVid. In the introductory implementation presented here, the values of the model parameters are chosen empirically. Although the same set of values works over a wide range of image types and application domains as demonstrated, making these parameters locally adaptive may improve the results or may be necessary in certain images and applications. From our observations, the expansion of the dynamic range at low illumination regions comes at the expense of saturating the bright regions. One example is shown in Fig. 10. In the Figure, the sky and cloud region lose some contrast compared to the same region in the original image. In fact, as shown in Fig. 8, the benchmark Zero-DCE algorithm performs slightly better in the cloudy regions of the images where the input image has high brightness.

As stated previously, the Gaussian spectral phase function was chosen for mathematical simplicity. Other phase functions, such 2D polynomials with adaptive parameters maybe investigated in future research. Another future direction is the hardware realization of the algorithm with diffractive optics and spatial light modulators



Fig. 10 Example of a limitation of the VEVid algorithm. Whereas dynamic range of the low intensity regions of the original image are extended, the higher intensity portion of the image within the clouds and sky become slightly saturated and lose some detail. VEVid Parameters: $S=0.3$, $b=0.39$, $G=0.8$

which will require interpreting the pixel brightness as the field squared, compared to the field in the present formulation.

Deep neural networks have proven powerful tools for object detection and tracking, and they are the key to autonomous driving and security systems, among others. We showed the utility of VEVid pre-processing to increase the accuracy of object detection by a popular neural network (YOLO). VEVid allows such neural networks that are trained on daylight images to generalize to night-time environments without having to be retrained. The application of VEVid to the color enhancement of digital images is also demonstrated.

5 Methods

For all results shown, computations were performed with a NVIDIA GeForce GTX TITAN X Graphic Processing Unit (GPU). The VEVid algorithm was built and run using PyTorch with support for asynchronous computation. Timing metrics were calculated using PyTorch's built-in asynchronous event objects. Zero-DCE's runtime results were computed using the code found at <https://github.com/Li-Chongyi/Zero-DCE>. This code has been made public by the authors of the original paper [17].

Image data from several common low-light image enhancement datasets, as images captured by us were used. Figures 1, 3 and 5 are from [24]. The IR camera security image (Fig. 6) is from [25]. Most of the images in Fig. 4 were taken with an iPhone 12 Pro Max, except the lighthouse image which is from [26]. The image for Fig. 7 is from [9] and [24]. The image in Fig. 9 is from a stock photo website [27]. The images in Figs. 11 and 12 are from [28]. Image in Fig. 10 is from [29].

The YOLOv3 object detection algorithm is used for benchmarking AI performance and is built using PyTorch with pretrained weights [30]. The computational speed results in Fig. 7 were obtained by averaging over a number of images. For each image, the frame size was varied by cropping.

5.1 Impact of model parameters

Here we describe the impact of the model parameters on the output image. Regarding the regularization term, b , in the numerical version of VEVid, the regularization is not fundamentally necessary but it does improve noise. In the closed-form mathematical approximation of VEVid described in Sect. 3, b is necessary. Without it, the real and imaginary components will be the same and spatial information will be lost.

The value of the S parameter is constrained by the requirement for small phase (nominal near field) diffraction and is usually in the 0.1–0.4 rad range. Figure 12 is obtained using the full numerical version of the algorithm. As shown in the left panel, when the regularization, b , is low, T has an appreciable impact. However, when regularization is high, it has less impact. This figure supports the approximation made for mathematical acceleration in which the T parameter does not appear because the spectral phase function is approximated as a constant, b .

The right panel in Fig. 12 shows the impact of b and G hyperparameters on the performance of VEVid. Increasing b reduces the effect of the imaginary part to which it is added. As can be seen in Fig. 11, being inverted, the imaginary part is a bright image, and reducing its influence has the effect of exposure reduction. Increasing b

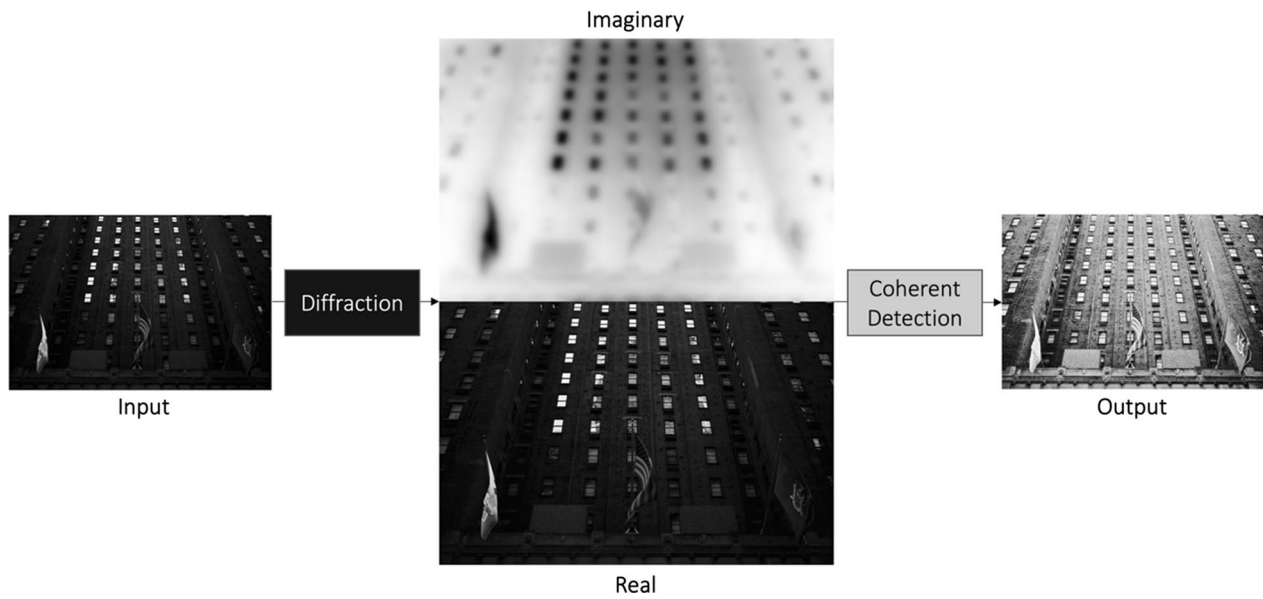


Fig. 11 An input image, and its real, imaginary, and phase components after virtual diffraction. We only show the V-channel because in low-light enhancement, VEVID is only applied to this channel

pushes the bright pixels into the saturating region of the arctan where they are squashed. The gain parameter, G , has a similar effect.

5.2 The intuition behind VEVID

This section of the paper has been added based on Reviewer’s request for the intuition behind VEVID and how it works. The intuition is not based on mainstream literature but rather has evolved from more than two

decades of research in our laboratory on time stretch instruments and the insights into dispersion, diffraction, nonlinearities and signal detection, gained therein. This information is added here in the Methods section in order not to overshadow the demonstrations of striking results produced of the algorithm. We emphasize that the physical intuition described in this section is for the full numerical version of VEVID which follows the physical analogies. It does not apply to the simplified closed form

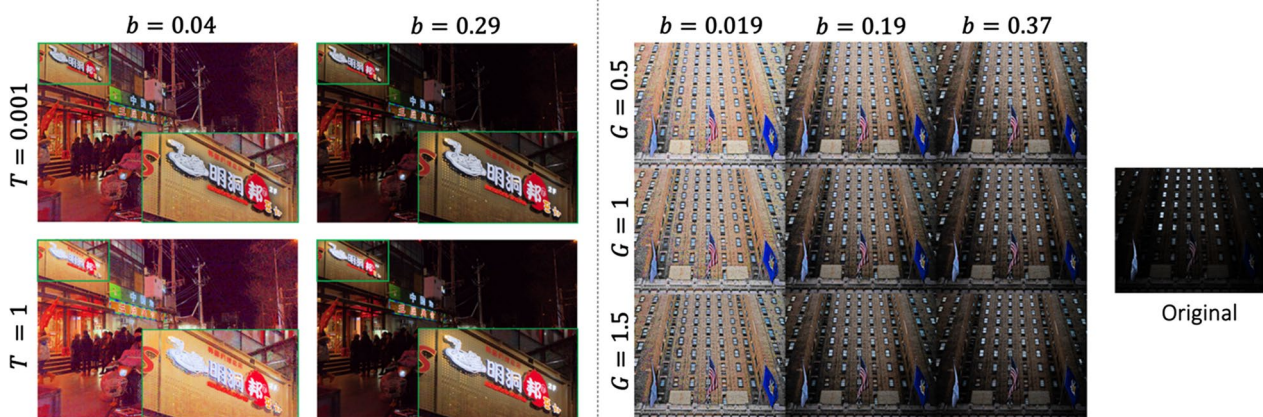


Fig. 12 Left panel: Impact of the regularization, b , and the variance of spectral distribution, T . When the regularization is low, T has an appreciable impact. However, when regularization is high, it has less impact. This figure supports the claims in the section on mathematical acceleration in which the T parameter is not needed. Right panel: The impact of hyperparameters on the performance of VEVID. Increasing b reduces the effect of the imaginary part to which it is added. As can be seen in Fig. 11, being inverted, the imaginary part is a bright image, and reducing its influence has the effect of exposure reduction. Increasing b pushes the bright pixels into the saturating region of the arctan where they are squashed. The gain parameter, G , has a similar effect

representation of VEViD because the approximations made to achieve the mathematical simplification render that model empirical; in other words non-physical.

First, we need to explain the general process of low-light enhancement. The enhancement requires a nontrivial nonlinear operation that enhances the dynamic range. It reshapes the energy distribution by making it more uniform—but without increasing the dynamic range (the number of quantization bits). This is the first requirement. The second requirement is that since this redistribution involves emphasizing the low-light (dark) pixels, the operation will increase noise as low-intensity pixels have poor signal to noise ratio. This must be avoided which means the nonlinear transformation must also be robust against noise.

VEViD performs low-light enhancement by reinterpreting a digital image as a discretized electromagnetic “lightfield” and subjects it to diffractive propagation over a short distance (small spectral phase) followed by coherent detection. The fact that this process equalizes and enhances the dynamic range of a signal making its features more vivid was first observed in our experiments on spectroscopy [31]. The 2014 paper was concerned with single-shot absorption spectroscopy enabled by the time stretch technique but with two unique features. First instead of utilizing far-field dispersion which requires an excessive amount of dispersion we utilized near-field dispersion which requires less dispersion (small spectral phase). Second, instead of detecting the brightness of the optical spectrum as is common practice in spectroscopy, we measured its phase through coherent detection. That paper provides the seed insight into how processes occurring in physical diffractive and coherent detection can provide a blueprint for designing digital algorithms that perform dynamic range enhancement. Figures 4 and 5 in [31] show that the phase of the dispersed pulse has a built-in equalization behavior that redistributes the weak and strong signals making the features in the signal more uniform, suppressing noise and enabling clear observation of weak features—properties that are needed for low-light enhancement. As stated in [31] “phase shifts [of the dispersed waveform] can be indicative of activity beyond the dynamic range of the amplitude measurements” [31].

Mathematically speaking, this equalization property can be seen by interpreting Eq. 9 of the present manuscript as follows. The division by the input image, $E_i[n, m; c]$, in the argument of the arctan function emphasize the low intensity regions of the image producing low-light enhancement. Subsequently, the arctan operation compresses the output preventing an undesirable dynamic range expansion and suppressing the noise.

Together these operations redistribute the energy while managing the dynamic range and noise.

The second insight behind VEViD is as follows. Why would an image have a phase in the first place? After all, an image is a real-valued vector with no imaginary component and hence has no phase. How is the image transformed from a real-valued vector to a complex-valued output image? As explained in our prior publications on the time stretch technique [32, 33], for 1-D temporal signals this occurs in the nominal near-field regime of temporal dispersion defined as the regime before the Stationary Phase Approximation is satisfied. For 2D images, the situation is the same as in 1D temporal signals when 1D time is replaced by 2D transversal coordinates (x, y) and 1D temporal frequency is replaced with 2D spatial frequencies (k_x, k_y). We note that “nominal near field” term introduced here is different from the textbook definition of near field diffraction which refers to the extreme near region where the solutions to the wave equation are exponentially decaying. In our “nominal near field” the solutions are propagating but the induced spectral phase is small.

According to Fourier Optics, physical diffraction is modeled by a Fourier transformation into the frequency domain, multiplication by a complex exponential with a frequency dependent phase exponent (the “propagator”), and inverse Fourier transformation back to the spatial domain. The “nominal near field” regime corresponds to a small propagation distance which is equivalent to a small phase value.

Under this condition, the output image is a complex-valued quantity, after it is transformed back into the spatial domain. This is also the insight behind the small value of the spectral phase used in VEViD. The small phase is equivalent to short propagation distance. It avoids phase wrapping and the resulting ambiguity. It also enables mathematical linearization the complex exponential propagator for computational simplification of the algorithm.

Mathematically, Eq. 6 of the manuscript describes the conversion of the real-valued input, $E[n, m; c]$, to a complex valued vector whose phase angle is used for the VEViD output, $V[n, m; c]$. This is achieved through multiplication of the image’s spectrum by the propagator, $e^{-i\phi[k_n, k_m]}$, followed by conversion back into the spatial domain.

The next key concept behind VEViD that needs to be explained is the shape of the spectral phase, i.e. its frequency dependence. In VEViD the image is transformed via a two-step process: (1) virtual diffraction applies a spectral phase function which converts the real-valued image into a complex-valued vector, and (2) coherent

detection computes the phase which has the desired transformation properties of equalization and noise tolerance. The spectral phase function emphasizes the phase of low frequencies and attenuates the phase of high frequencies. The best way to understand how the spectral phase affects the image in the spatial domain can be understood using the Fourier integration property. To gain the basic insights, consider a very simple spectral phase function, $\phi[k_n, k_m]$, that is inversely proportional to the frequency, k_n, k_m . It leads an integration operation in the spatial domain and has useful properties of compressing the extreme variations in pixel intensities as well as averaging and noise reduction.

As seen in Eqs. 10 and 11, in the small phase approximation, the complex exponential phase propagator is linearized leading to the phase appearing only in the imaginary component of the diffracted signal. This effect is clearly visible in the imaginary component shown in Fig. 11. The figure shows an input image, and its real, imaginary, and phase components after virtual diffraction. We only show the V-channel because in low-light enhancement, VEViD is only applied to this channel, as explained in the paper. This figure clearly shows that the real part is nearly the same as the input, the imaginary part follows properties consistent with integration and the phase has the desired low-light enhancement. An interesting complimentary behavior between the real and imaginary components is also observed.

Abbreviations

c	Color channel
$E(x, y; c)$	"Lightfield" in continuous spatial coordinates
$\tilde{E}(k_x, k_y; c)$	"Lightfield" in continuous spatial-frequency coordinates
$[n, m]$	Spatial discrete (pixelated) coordinates
$[k_n, k_m]$	Spatial-frequency discrete (pixelated) coordinates
$E_i[n, m; c]$	Input image
$E_o[n, m; c]$	Output after propagation
V	Output Image (VEViD transform)
$\phi[k_n, k_m]$	Spectral phase function, $\phi[k_n, k_m] = S \cdot \hat{\phi}[k_n, k_m]$
$H[k_n, k_m]$	Propagation operator (propagator), $H[k_n, k_m] = e^{-i\phi[k_n, k_m]}$
Phase Variance(T)	Variance of the spectral phase function $\phi[k_n, k_m] = \exp\left[-\frac{k_n^2 + k_m^2}{T}\right]$
Phase Scale(S)	Phase strength $\max\phi[k_n, k_m] = \max(S \cdot \hat{\phi}[k_n, k_m])$
Phase Activation Gain (G)	Phase activation gain, $V = \tan^{-1}\left(G \cdot \left(\frac{\Re\{E_o\}}{\Im\{E_o\}}\right)\right)$
b	Regularization constant
$N\{\}$	Normalization function

Acknowledgements

The authors thanks Yiming Zhou in Jalali Lab for helpful discussions.

Author contributions

BJ conceived the general concept and directed the research. CM performed the coding and contributed to refining the idea. Both authors performed the

analytics and wrote the manuscript. Both authors read and approved the final manuscript.

Funding

This work was partially supported by the Parker Center for Cancer Immunotherapy (PICI), Grant No. 20163828, and by the Office of Naval Research (ONR) Multi-disciplinary University Research Initiatives (MURI) program on Optical Computing Award Number N00014-14-1-0505.

Availability of data and materials

Codes and data will be made available on Github once the paper is published.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Bahram Jalali serves as an Editor for the journal, no other author has reported any competing interest.

Received: 26 August 2022 Revised: 27 September 2022 Accepted: 29 September 2022

Published: 8 November 2022

References

1. https://en.wikipedia.org/wiki/Antikythera_mechanism
2. Donald Routledge Hill, *Mechanical Engineering in the Medieval Near East*, Scientific American, 1991, pp. 64–69 (cf. Donald Routledge Hill, *Mechanical Engineering*)
3. D.R. Solli, B. Jalali, Analog optical computing. *Nat. Photonics* **9**(11), 704–706 (2015)
4. D.R. Solli, C. Ropers, P. Koonath, B. Jalali, Optical rogue waves. *Nature* **450**(7172), 1054–1057 (2007)
5. J.M. Dudley, G. Genty, A. Mussot, A. Chabchoub, F. Dias, Rogue waves and analogies in optics and oceanography. *Nat. Rev. Phys.* **1**(11), 675–689 (2019)
6. E.H. Land, J.J. McCann, Lightness and Retinex theory. *J. Opt. Soc. Amer.* **61**(1), 1–11 (1971)
7. E.H. Land, The retinex theory of color vision. *Sci. Am.* **237**, 108–129 (1977)
8. M. Li et al., Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **27**(6), 2828–2841 (2018)
9. X. Guo, Y. Li, H. Ling, 'LIME: low-light image enhancement via illumination map estimation.' *IEEE Trans. Image Process.* **26**(2), 982–993 (2017)
10. L. Li, S. Sun, C. Xia, Survey of histogram equalization technology. *Comput. Syst. Appl.* **23**(3), 1–8 (2014). (in Chinese)
11. G. Yadav, S. Maheshwari, A. Agarwal, *Contrast limited adaptive histogram equalization based enhancement for real time video system*, 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 2392–2397, <https://doi.org/10.1109/ICACCI.2014.6968381>.
12. K.G. Lore, A. Akintayo, S. Sarkar, LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit* **61**, 650–662 (2017)
13. Lv F, Lu F, Wu J, Lim C. (2022). MBLLEN: Low-light Image/Video Enhancement Using CNNs.
14. M. Zhu, P. Pan, W. Chen, Y. Yang, EEMEFN: low-light image enhancement via edge-enhanced multi-exposure fusion network. *Proc. AAAI Conf. Artif. Intell.* **34**(07), 13106–13113 (2020). <https://doi.org/10.1609/aaai.v34i07.7013>
15. K. Lu, L. Zhang, TBEFN: a two-branch exposure-fusion network for low-light image enhancement. *IEEE Trans. Multimedia* **23**, 4093–4105 (2021). <https://doi.org/10.1109/TMM.2020.3037526>

16. Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, Enlightengan: deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **30**, 2340–2349 (2021)
17. C. Guo, C. Li, J. Guo, C. Loy, J. Hou, S. Kwong, R. Cong (2020). Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1780–1789)
18. M. Asghari, B. Jalali, Edge detection in digital images using dispersive phase stretch transform. *Int. J. Biomed. Imaging* **2015**, 687819 (2015)
19. M. Suthar, B. Jalali, "Phase-stretch adaptive gradient-field extractor (PAGE)," in *Coding Theory*, 2020.
20. C. Neil. Thompson et al. (MIT) 2020 <https://arxiv.org/pdf/2007.05558.pdf>
21. J Redmon, A. Farhadi (2018). YOLOv3: An Incremental Improvement. *ArXiv*, abs/1804.02767.
22. C. Li, C. Guo, L.H. Han, J. Jiang, M.M. Cheng, J. Gu, C.C. Loy, Low-light image and video enhancement using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **01**, 1–1 (2021)
23. B. Jalali, *Device Physics & Modeling*, In: *InP HBTs: Growth, Processing, and Applications*, Jalali, B., Pearson, S.J. (Eds.), (0-89006-724-4). PP. 229 - 263 (December 1994) Artech House
24. W. Yang, Y. Yuan, W. Ren, J. Liu, W. Scheirer, Z. Wang, T. Zhang et al., Advancing image understanding in poor visibility environments: a collective benchmark study. *IEEE Trans. Image Process.* **29**, 5737–5752 (2020)
25. X. Jia, et al. *LLVIP: A visible-infrared paired dataset for low-light vision*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
26. K. Ma et al., perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* **24**(11), 3345–3356 (2015)
27. <https://www.pexels.com/video/neo-classical-building-in-city-at-night-9935075/>.
28. <https://photographylife.com/underexposure-and-overexposure-in-photography>
29. <https://sites.google.com/site/vonikakis/datasets/tm-died>
30. <https://pjreddie.com/media/files/yolov3.weights>.
31. P.T. DeVore, B.W. Buckley, M.H. Asghari, D.R. Solli, B. Jalali, Coherent time-stretch transform for near-field spectroscopy. *IEEE Photonics J.* **6**(2), 1–7 (2014)
32. D.R. Solli, S. Gupta, B. Jalali, Optical phase recovery in the dispersive Fourier transform. *Appl. Phys. Lett.* **95**(23), 231108 (2009)
33. K. Goda, D.R. Solli, K.K. Tsia, B. Jalali, Theory of amplified dispersive Fourier transformation. *Phys. Rev. A* **80**(4), 043821 (2009)