

# Turn a Silicon Camera into an InGaAs Camera

Feifan Lv<sup>1</sup>

Yinqiang Zheng<sup>2</sup>

Bohan Zhang<sup>1</sup>

Feng Lu<sup>1,3,4,\*</sup>

<sup>1</sup>State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University, Beijing, China

<sup>2</sup>National Institute of Informatics, Japan

<sup>3</sup>Peng Cheng Laboratory, Shenzhen, China

<sup>4</sup>Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, China

## Abstract

Short-wave infrared (SWIR) imaging has a wide range of applications for both industry and civilian. However, the InGaAs sensors commonly used for SWIR imaging suffer from a variety of drawbacks, including high price, low resolution, unstable quality, and so on. In this paper, we propose a novel solution for SWIR imaging using a common Silicon sensor, which has cheaper price, higher resolution and better technical maturity compared with the specialized InGaAs sensor. Our key idea is to approximate the response of the InGaAs sensor by exploiting the largely ignored sensitivity of a Silicon sensor, weak as it is, in the SWIR range. To this end, we build a multi-channel optical system to collect a new SWIR dataset and present a physically meaningful three-stage image processing algorithm on the basis of CNN. Both qualitative and quantitative experiments show promising experimental results, which demonstrate the effectiveness of the proposed method.

## 1. Introduction

Human eyes can perceive a small portion of light only, usually lying in the wavelength range of 400-700nm. To mimic the trichromatic color perception of human eyes, existing digital cameras are usually equipped with a Silicon sensor, together with a color filter array to filter out light beyond the visible spectrum (VIS). However, a typical Silicon sensor itself is highly sensitive to light in a wide range of 300-950nm. This property has allowed numerous applications of Silicon based cameras in the ultraviolet (UV) range less than 400nm and the near infrared (NIR) range beyond 700nm. For example, security surveillance cameras utilize the sensitivity of a Silicon sensor around 940nm, so as to see objects clearly illuminated by NIR LED lamps.

Short-wave infrared (SWIR) lies between NIR and mid-

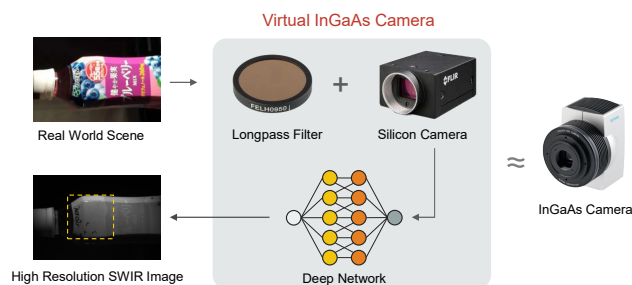


Figure 1. The schematic diagram of our proposed solution for short-wave infrared imaging using a Silicon sensor. Our solution combines a Silicon camera, a longpass filter and a CNN based image processing algorithm to build a virtual InGaAs camera.

wavelength infrared (MWIR). It is less sensitive to thermal irradiation than MWIR, and thus allows high quality imaging with sufficient contrast, similar to VIS and NIR light. On the other hand, the relatively longer wavelength of SWIR leads to stronger penetration capability, which is proven crucial in industrial inspection and visibility enhancement under extreme weather condition. For instance, the defects inside a nice-looking apple are invisible to our eyes, yet become apparent when viewed by a SWIR imaging device, and objects hidden in the heavy fog are likely to be seen in the SWIR range.

Capturing SWIR images requires special sensors. In contrast to those alternatives like HgCdTe and InSb, the InGaAs sensor is most commonly used since it can work stably at the room temperature, and it also has the advantages of relatively low power, small volume, high sensitivity, etc.. In spite of these, when compared with the Silicon sensor, the InGaAs sensor suffer from a variety of drawbacks, including low spatial resolution, high price and high pixel defection rate, which severely limit the widespread use of InGaAs in SWIR imaging.

Efforts have been taken to resolve those drawbacks. For example, the compressive sensing techniques have been utilized to boost the spatial resolution of a low-resolution InGaAs camera [6], or even construct an area SWIR camera

\*Corresponding Author: Feng Lu (lufeng@buaa.edu.cn)

This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grant 61602020 and Grant 61732016.

by using a single-pixel photo detector [8]. However, the reconstruction of such an SWIR imaging system is non-trivial, due to the involvement of a digital micromirror device (DMD) for coded sampling. It requires multiple scans, which prevents its application for dynamic scenes, and the aforementioned drawbacks of InGaAs sensors can not be completely avoided, since at least one low-end InGaAs detector has to be used.

Different from all existing solutions, in this paper we propose to replace the InGaAs sensor by a common Silicon sensor for snapshot SWIR imaging, as illustrated in Figure 2. The required hardware is an ordinary Silicon camera, mounted with a 950nm long pass filter. A high-resolution image, which approximates the SWIR image captured by an InGaAs camera, will be generated by our trained end-to-end deep neural network. As a result, our system has the advantages of higher sensor maturity, lower price and higher imaging resolution. However, this is an extremely challenging task, since the sensitivity of a Silicon sensor is extremely weak in the SWIR range. Through experiments, we have found that a Silicon sensor can perceive light up to 1200nm, yet the typical response range of an InGaAs sensor is between 950nm and 1650nm.

Our key idea is to properly utilize the sensitivity of a Silicon sensor, weak as it is, in the 950-1200nm range, so as to approximate the response of an InGaAs sensor. To this purpose, we develop a six-channel optimal imaging system to collect a novel SWIR dataset, which allows to train a physically interpretable three-stage Silicon-to-InGaAs network mapping. Both qualitative and quantitative experiments have shown that our proposed three-stage method is capable of generating realistic SWIR images, which outperforms the direct image-to-image mapping counterparts. Overall, our major contributions are threefold:

- We propose a bold and novel solution for snapshot, low-cost and high-resolution SWIR imaging by using a Silicon sensor, rather than an InGaAs sensor.
- We design a six-channel imaging system to acquire data and build a new dataset with six paired Silicon/InGaAs images, which allows Silicon-to-InGaAs mapping in a more principled way.
- We design a novel end-to-end three-stage image processing algorithm for our solution, which is physically interpretable from the perspective of multispectral imaging.

To the best of our knowledge, this is the first attempt in the short-wave infrared imaging field.

## 2. Related Work

Since SWIR imaging shows a great potential in a wide range of applications, plenty of researches have been re-

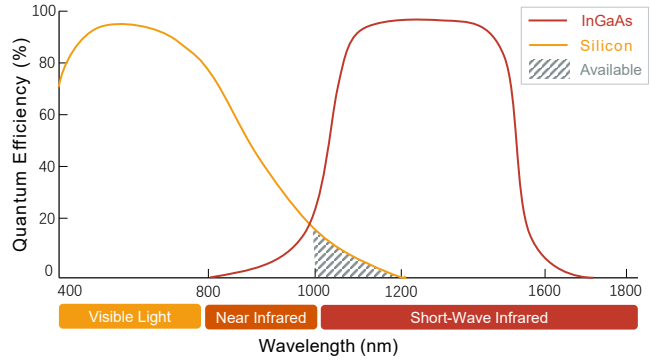


Figure 2. The typical Quantum Efficiency (QE) curve of Silicon and InGaAs sensors. It can be seen that the Silicon has poor QE for the short-wave infrared region, however, InGaAs has excellent QE in that region. Our goal is to simulate the signal of InGaAs cameras according to the slash area which represents the data in the short-wave infrared region acquired by Silicon sensors.

ported recently and most of them focus on either image quality improvement or applying it to specific applications. For the former, recent works can be further divided into the compressive sensing (CS)-based and the image enhancement technology-based. In this section, we overview all these categories as related works.

Notice that to the best of our knowledge, there is no previous work using a Silicon camera for SWIR imaging, making our work have *no directly related work*.

**Compressive sensing.** Compressive sensing (CS) [6] deals with the problem of recovering a signal from far fewer samples than those fulfilling the sampling theorem. Although this is an ill-posed problem, robust estimation can be done under some conditional constraints [2, 26]. The main benefit of using CS is reducing the hardware requirement for image resolution, especially in the case of SWIR or Terahertz (X-ray). A typical example is the single pixel camera (SPC) [8] which uses digital micromirror devices (DMD) and CS algorithm for image reconstruction with only one single sensor. Ke *et al.* [17] present a block-based compressive imaging system to reach superior reconstruction performance. Kerviche *et al.* [18] present a scalable compressive imager using information-optimal measurements design and single-pass piece-wise linear reconstruction algorithm. Mahalanobis *et al.* [24] demonstrates that high spatial resolution information in the mid-wave infrared can be successfully recovered by small focal plane arrays measurements. LiSens [30] replaces the photodetector in the SPC with a 1D linear array or a line-sensor to improve the measurement rate. FPA-CS [3] proposes a focal plane array-based compressive sensing architecture using an array of SPCs in parallel to increase the measurement rate. Besides, there are many researches that combines various signal models and constraints to improve the quality of video

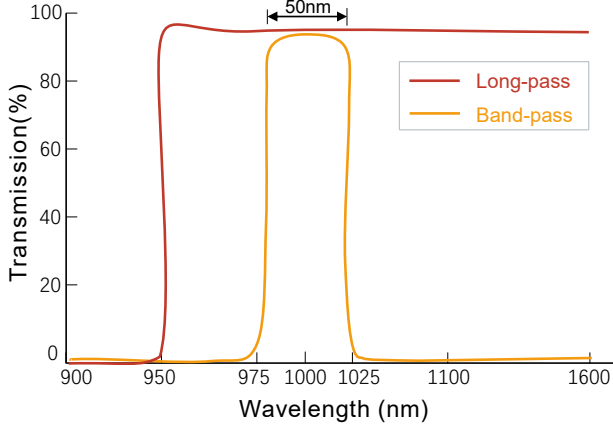


Figure 3. Transmission curve of the 950 nm longpass filter and the 1000 nm bandpass filter. The full width-half max of the bandpass filters in this research is 50 nm. By combining the filters flexibly, we are able to capture the signal of specific spectral regions.

imaging, like optical flow-based reconstructions [29], etc.

Although the above CS-based techniques can reduce the dependance on hardware resolution for imaging, at least one (low-resolution) expensive SWIR sensor (like InGaAs sensor) is still needed. Besides, such technique can only handle static scenes. For dynamic scenes, the CS-based methods can help increase the frame rate of the video [21, 10, 13] but it cannot help improve the resolution.

**Image enhancement.** Image processing technology, such as super-resolution (SR), is an almost zero-cost way to improve the quality of SWIR imaging. With the development of deep learning, many super-resolution algorithms [20, 16, 27] can be used to enhance the InGaAs images. However, this requires large-scale SWIR image datasets with enough detail information, which by now, has not been reported yet. Dong *et al.* [5] implement spatially adaptive iterative singular-value thresholding(SAIST) to restoring SWIR images and manage to significantly suppress the noise level. Nevertheless, even with such technique, at least one InGaAs camera is still needed.

**Applications.** Owing to its unique characteristics, SWIR shows great potential in a wide range of applications that are difficult or even impossible to perform in the visible spectrum, for instance, agricultural and industrial product inspection, process quality control, surveillance and anti-counterfeiting, biological imaging technique, Optical Coherence Tomography imaging and much more [3, 11]. SWIR imaging has already been playing an important role in observation under hazy weather conditions and low-light conditions [9]. There has also been study showing that SWIR spectroscopy is very useful for measuring mineral physicochemistry sensitive to changes in metamorphic grade, especially in very low-grade rocks [7].

### 3. Proposed Solution

As described above, conventional SWIR imaging researches rely on specific sensors. Meanwhile, due to the limitations in materials science and craftsmanship, the price of SWIR sensors will not drop significantly in the short term. Therefore, these existing researches still require high cost hardware. In this paper, we propose a novel solution that can avoid using expensive SWIR sensors. The comparison with some existing researches is shown in Figure 4.

#### 3.1. Optical Design

Firstly, we need to know why Silicon sensors can be used for SWIR imaging. Figure 2 shows the typical quantum efficiency curve of Silicon and InGaAs sensors, where the InGaAs sensor is a typical type of SWIR sensors. Since the Silicon sensor is not sensitive to short-wave infrared, it has a very low quantum efficiency in SWIR area which is also the main reason why it cannot be directly used for SWIR imaging. Another reason is that the high sensitivity of the Silicon sensor to the visible light can seriously affect the imaging in SWIR area.

The latter problem can be simply solved by using some optical filters to only keep the desired spectrum band. Following this idea, in this paper we choose to use a longpass filter and a set of bandpass filters, as shown in Figure 3. By combining the filters flexibly, we are able to suppress the effects of visible light for the Silicon sensor and only capture light within certain wavelengths.

For the problem that the Silicon sensor is not sensitive to SWIR, it is difficult to reconstruct the SWIR images directly from the Silicon images without additional information. On the other hand, it can be much easier to learn individual mappings from a Silicon sensor to a SWIR sensor within each narrower wavelength band. Based on this consideration, we design a novel pipeline for SWIR imaging using a Silicon sensor as shown in Figure 5. We use bandpass filters to produce multi-channel signals with respect to different wavelengths, and try to simulate the responses of all these channels to approximate the InGaAs image.

Besides, we also introduce the additional light sources to

Name	Resolution	Video	Price [3]	Hardware
Nyquist Sampling	low	yes	\$\$\$\$\$	High res. Sensor array
SPC [8]	high	limited	\$\$\$	DMD + photodetector
CS-MUVI [29]	medium	yes	\$\$\$	DMD + photodetector
FPA-CS [3]	high	yes	\$\$\$	DMD + Low res. Sensor array
Ours	higher	yes	\$	Silicon Sensor array + Long pass Filter

Figure 4. The comparison of imaging resolution, dynamic scene support, manufacturing cost and hardware requirements against some existing approaches.

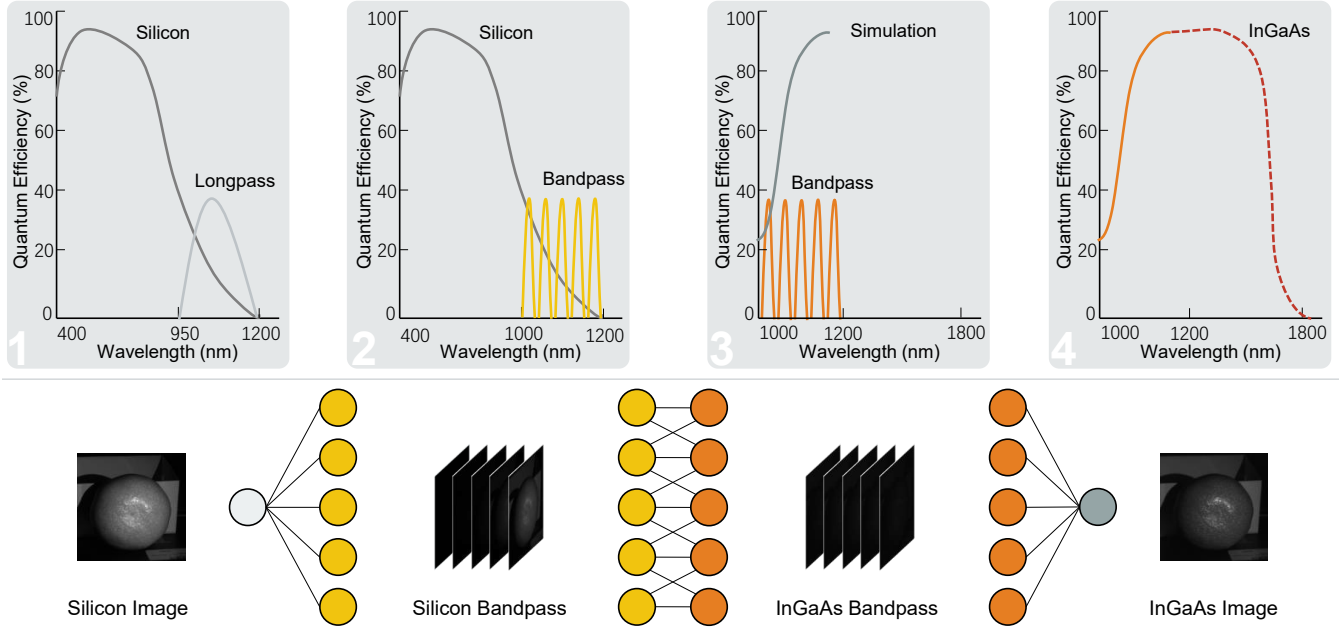


Figure 5. The schematic diagram of short-wave infrared imaging using Silicon sensors. The lower part is the schematic diagram of the SWIR imaging pipeline process, and the upper part is the optical design and theoretical analysis diagram corresponding to each part result. The detailed description can be found in the text.

solve the low signal-to-noise ratio problem caused by insensitivity. This is easy to apply in many application scenarios.

### 3.2. Proposed SWIR Imaging Pipeline

As shown in Figure 5, the SWIR imaging pipeline consists of four steps: acquisition, decomposition, simulation, and reconstruction.

1. We capture Silicon images as the input using a Silicon sensor with a 950 nm longpass filter which is used to shield visible light.

2. The input image is decomposed into a set of images, each of which represents the signal of narrower wavelength band. This process can be viewed as physically using bandpass filters to intercept particular wavelength bands signal from the longpass filter signal.

3. Different mapping relationships for each wavelength band are learned to simulate the InGaAs signal using Silicon signal.

4. The InGaAs image is reconstructed based on the simulated signal.

It is worthy of mentioning that Silicon signal does not cover all SWIR areas, as shown in Figure 2. The reconstruction process is similar to using sampled signal (simulated InGaAs bandpass images) to recover the overall signal (InGaAs longpass image).

Overall, our pipeline is a solution that combines software (image processing algorithm) and hardware (Silicon sensor with a longpass filter). Image processing can be done in a

variety of ways and our designed network is a specifically designed one. The detail architecture can be found below.

## 4. Hardware & Dataset

In order to support the proposed approach, we design a multi-channel imaging system and collect a new dataset for training and benchmarking. To the best of our knowledge, this is the first public dataset for SWIR imaging using Silicon cameras.

### 4.1. Hardware Configuration

The design schematic illustration and prototype of the imaging system are shown in Figure 6. The resolution of the used Silicon camera (GS3-U3-15S5M-C) is  $1384 \times 1032$  and the pixel depth is 14 bit. Unlike high-resolution Silicon cameras, the resolution of used InGaAs camera (BK-51IGA) is  $128 \times 128$  and it needs to be cooled to  $-5^\circ\text{C}$  to work properly. The pixel depth of InGaAs image is 16 bit.

We use a motorized rotator (Thorlabs FW102C) for automatic filter switching to ensure that every set of images taken by each camera is aligned. It has 6 holes, which are used to separately place 950 nm longpass filter (Thorlabs FELH0950) and five bandpass filters (1000 nm, 1050 nm, 1100 nm, 1150 nm and 1200 nm CWL, Edmund Hard Coated OD 4 50 nm bandpass Filter). The Full Width-Half Max (FWHM) of all bandpass filters is 50 nm. In order to ensure the physical alignment of the imaging, we use a



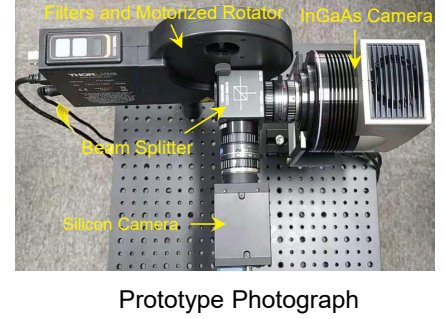
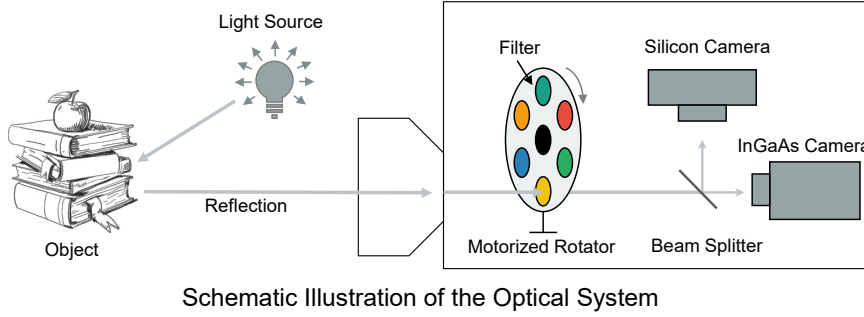


Figure 6. The hardware implementation of our designed multi-channel imaging system. The system consists of a light source, an InGaAs camera, a Silicon camera, a beam splitter and a motorized rotator with six different filters (one longpass filter and five bandpass filters).

beam splitter (Thorlabs CCM1-BS015) whose splitting ratio is 50 : 50 and coverage wavelength range is from 900 *nm* to 1600 *nm*. The light source we used is a halogen lamp. In addition, we kept the camera settings and position fixed throughout the imaging process.

#### 4.2. Silicon-to-InGaAs Dataset

As mentioned above, the imaging pipeline consists of four steps: acquisition, decomposition, simulation, and reconstruction. Among them, the image obtained in the first step is used as the input, and each of the other steps needs to be given a ground truth to constrain the input. This means that for every scene, in addition to the input Silicon image and the corresponding InGaAs image, we need to collect another 10 images to guide the second and third steps. That is to say, we should collect 12 images (2 cameras  $\times$  6 filters) for every scene.

Great efforts have been made to align the images physically, however, we still need to use alignment algorithm to adjust them because of the vast difference between the two

cameras in many aspects like resolution.

Due to the low brightness of the images, some well-known alignment algorithms such as Lucas-Kanade [23], SIFT [22] and their various deformations cannot achieve perfect results. Therefore, we use the method of manually selecting feature points to align the images, as shown in Figure 7. After alignment and cropping, the resolution of grayscale images in the dataset is  $128 \times 128$  and the bit depth of every image is 16 *bit*.

We collect 1560 images on 130 scenes to ensure the diversity of data. In this paper, we use the first 70 scenes as the training set and the remaining 60 scenes as the test set.

### 5. Methodology

A specific implementation of the proposed solution is introduced in this section with all the necessary details. As the analysis above, our solution focus on three major processes: decomposition of longpass images, InGaAs bandpass images simulation and reconstruction of InGaAs image based on the simulation results. We design a new network architecture consists of three sub-networks corresponding to the three processes as the image processing algorithm in our solution.

#### 5.1. Network Architecture

As shown in Figure 8, the proposed network consists of three sub-networks: Decomposition-Net, Simulation-Net and Reconstruction-Net. We use U-Net [28, 15] and Res-Net [12, 20] as the basic element of our network since they have been proven effective extensively. We removed Batch-Normalization [14] layers from the original implementation because test results show that they will slightly reduce the performance in our specific task.

For the U-Net module, we use convolutional layers with stride 2 for down-sampling and use resize-convolutional layers for up-sampling, which will suppress the checkerboard artifacts to some extent [25]. As the size of the feature map decreases, the number of feature maps doubles and the feature map number of the first layer is set to 32.

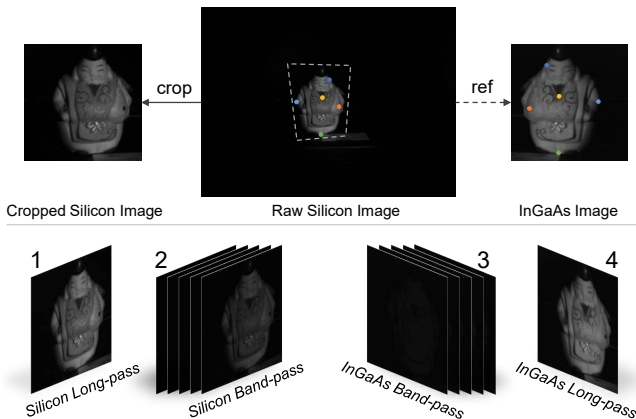


Figure 7. The upper part is a schematic diagram of the alignment and crop operation for the original Silicon images. The lower part is a sample in the SI dataset and each sample in the dataset contains 12 images. Objects captured in this dataset are diverse.

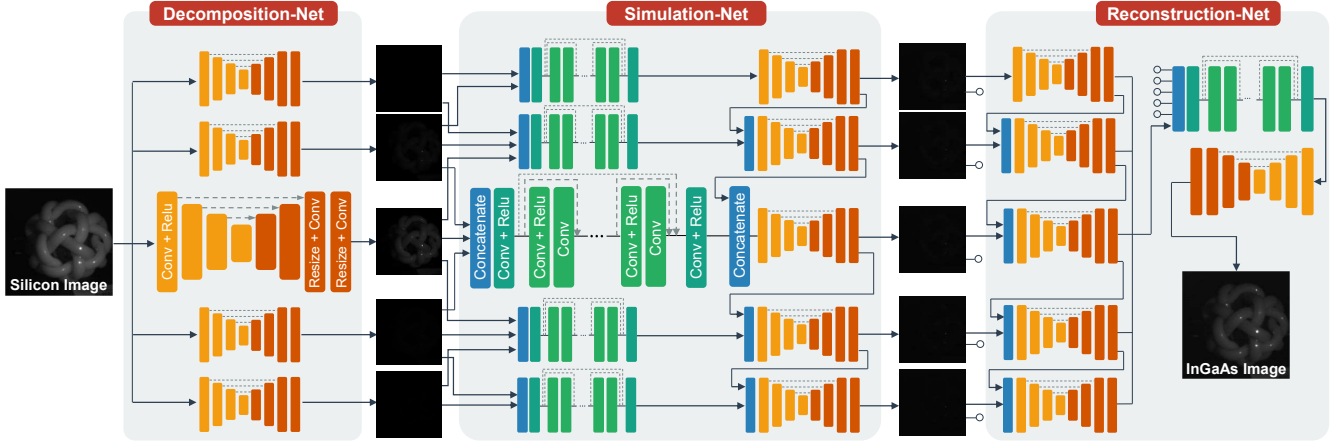


Figure 8. The proposed network architecture. It contains three subnetworks: Decomposition-Net, Simulation-Net and Reconstruction-Net, which correspond to the three network in Figure 5. The dashed lines represent skip connections, the circles represent discontinuous connections, the same color represents the same structure.

The Res-Net module consists of a convolutional layer, three resblocks, and a convolutional layer sequence. For every layer of the Res-Net module, the size of feature maps remains the same as the input and the number of feature maps for all layer is set to 32.

For all layers of both modules, we use the Rectified Linear Unit (ReLU) as the activation function and the kernel size is set to  $3 \times 3$ . Concatenating along the channel dimension instead of adding directly is used for skip connections.

**Decomposition-Net.** Since it requires five different physical bandpass filters to decompose the longpass Silicon image into five different bandpass images, we use five U-Net models as virtual bandpass filters to simulate the real filters. Therefore, the Decomposition-Net contains five branches and each branch is composed of one U-Net module. The input is the longpass Silicon images and the outputs are the simulation of five different bandpass Silicon images.

**Simulation-Net.** We design the Simulation-Net to simulate the InGaAs bandpass images according to the corresponding Silicon bandpass images. The input is the outputs of Decomposition-Net and the outputs are the simulation of five different InGaAs bandpass images. Considering that adjacent bands have similar properties, we also use the estimate of the adjacent bands as the input of the current band. With the wavelength increasing, the quantum efficiency of the Silicon sensor gradually decreases as shown in Figure 2. So we use the results of shorter wavelength regions to guide the simulating process of the adjacent longer regions.

**Reconstruction-Net.** In order to reconstruct the InGaAs image over the whole SWIR spectrum based on the obtained simulated InGaAs bandpass results, we design the Reconstruction-Net. In addition to directly using them as input, we also use U-Net modules to extract features from

the outputs of Simulation-Net as part of the input. Besides, we adopt a guiding strategy similar to the Simulation-Net.

## 5.2. Loss Function

According to our solution, we not only need to constrain the final result, but also constrain the intermediate results. That's why we need to give different weights to different sub-networks. At the same time, sub-networks with multiple outputs also require different weights to constrain their outputs. The total loss function is formulated as:

$$Loss = \alpha \cdot L_{Decom} + \beta \cdot L_{Sim} + \gamma \cdot L_{Recon}, \quad (1)$$

where the  $L_{Decom}$ ,  $L_{Sim}$  and  $L_{Recon}$  represent the loss function of Decomposition-Net, Simulation-Net and Reconstruction-Net, and  $\alpha$ ,  $\beta$ ,  $\gamma$  are the corresponding coefficients. The details of three loss are given below.

Due to the low brightness of the image, only using common error metrics such as MSE or MAE may cause blurring. To maintain structural consistency, we use the well-known image quality assessment algorithms SSIM [31] along with MAE as the loss function. The value ranges of SSIM is  $(-1, 1]$ . The definition of SSIM can be found in [31]. The  $L_{Decom}$  is defined as:

$$L_{Decom} = \sum_{i=1}^5 \lambda_i \cdot (L_{mae}^i + 1 - L_{ssim}^i), \quad (2)$$

where  $L_{mae}^i$ ,  $L_{ssim}^i$ ,  $\lambda_i$  represents the MAE, SSIM value and coefficient of the  $i$ -th output image. We use the same loss function for Simulation-Net,

$$L_{Sim} = \sum_{i=1}^5 \varphi_i \cdot (L_{mae}^i + 1 - L_{ssim}^i), \quad (3)$$

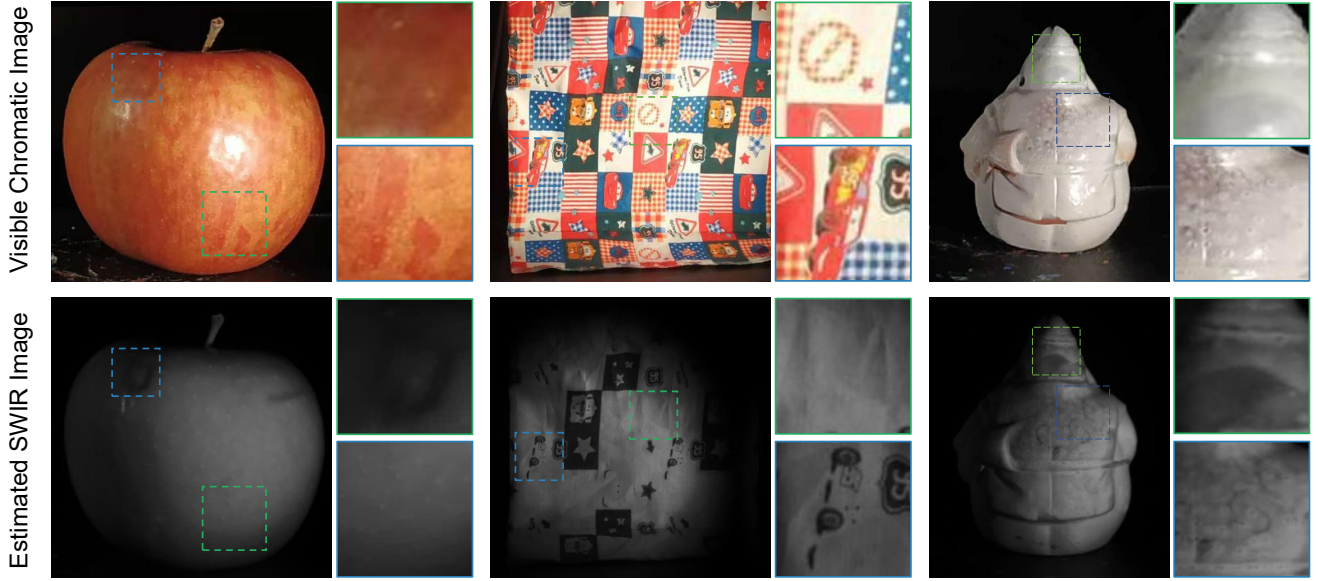


Figure 9. Some of the results from the proposed solution for short-wave infrared imaging. The upper part is the chromatic scenes while the lower part shows the estimated corresponding SWIR image. We amplify the framed part for better clarity.

where  $\varphi_i$  represents the coefficient of  $i$ -th output image and the other parameters have the same meaning as formula 2. The  $L_{Recon}$  is similarly defined as:

$$L_{Recon} = L_{mae} + 1 - L_{ssim}, \quad (4)$$

where  $L_{mae}$  and  $L_{ssim}$  have the same meanings as before. In the experiments, the configuration of the parameters is:  $\alpha = 100$ ,  $\beta = 10$ ,  $\gamma = 1$ ,  $\lambda_{1-5} = (1, 2, 5, 5, 5)$  and  $\varphi_{1-5} = (1, 2, 5, 5, 5)$ .

### 5.3. Implementation Details

Our implementation is done with Keras [4] and Tensorflow [1]. The proposed network can be quickly converged after being trained for 6000 epochs on a Titan-X GPU using the proposed dataset. In order to prevent overfitting on the small amount of data, we use random clipping, flipping and rotating for data augmentation. We set the batch-size to 30 and the size of random clipping patches to  $80 \times 80 \times 1$ . The input image values is scaled to  $[0, 1]$ .

In the experiment, training is done using the Adam optimizer [19] with  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10^{-8}$ . We also use the learning rate decay strategy, which reduces the learning rate to 95% before the next epoch. At the same time, we reduce learning rate to 50% when the loss metric has stopped improving.

## 6. Experiments

To demonstrate the performance of our solution, we first show results on several static scenes to highlight the capabilities of SWIR imaging. Then, we evaluate the effectiveness by comparing with several typical end-to-end network.

### 6.1. SWIR Imaging

We present a simple experiment to prove the feasibility of our solution. Figure 9 visually highlights the disparity of the scene captured in visible light spectrum and SWIR. When viewed in visible light, defect of the apple in the first set of images is perfectly hidden behind the overt texture, meanwhile, the pattern of the pottery figurine in the third image set is rather blurry. However, we can clearly identify them in the estimated SWIR image. Moreover, in the second scene, the estimated image captures the real shading of the surface without being affected by different colors. These results provide us with a glance at the potential applications of our system like agricultural and industrial product inspection, process quality control and so on.

### 6.2. Comparative Experiments

Since this is the first attempt to image SWIR with a Silicon camera to the best of our knowledge, there is not any prior results to compare with. In order to prove the effectiveness of our approach, we reimplemented some end-to-end networks that have been proven effective in image processing field to address our problem. Learning the mapping from Silicon to InGaAs directly is challenging, that's why we divide the problem into several simpler sub-problems. With the leverage of multi-channel information, our solution achieves superior outcome. We use the commonly-used SSIM [31] and PSNR as the metrics for evaluation as shown in Table 1. Visualization can be found in Figure 10, from which we can perceive that our results show the highest similarity with the InGaAs images.

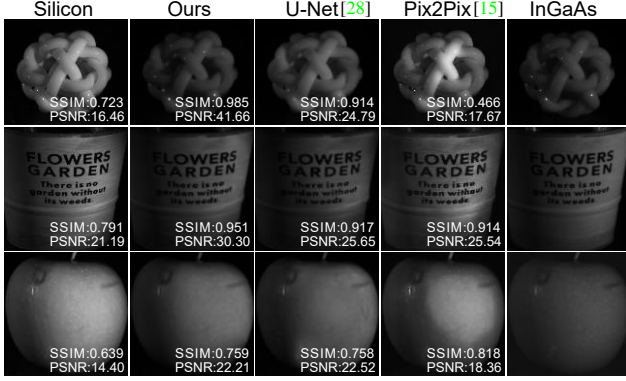


Figure 10. Visualization of the comparative experiments. Our solution is not only more similar to InGaAs image, but also has higher objective evaluation metrics.

	Ours	U-Net [28]	CGAN [15]
PSNR	<b>31.06</b>	28.20	20.34
SSIM [31]	<b>0.869</b>	0.855	0.644

Table 1. Comparative experiment results against two famous network architectures. We reimplemented the networks according to the original ones due to the particularity of our problem. This table reports mean PSNR and SSIM on the test set. Our approach leads in both metrics.

### 6.3. Bandpass Analysis

In this subsection, we apply bandpass analysis on the images in our dataset. In particular, we compute the PSNRs between images of different bandpasses, as shown in the two Confusion Matrices in Figure 11. We can see that bandpass images captured by the Silicon sensor differ significantly from each other, while those by the InGaAs sensor are more alike. This indicates that our dataset correctly capture the real property of the quantum effectiveness curves of different sensors as shown in Figure 2.

We specifically design an experiment to demonstrate the interpretability and effectiveness of the intermediate results. As is shown in Figure 12, the outputs of the Decomposition-Net are very similar visually to the signal intercepted by the real filters in each wavelength, which can also be reflected quantitatively from the PSNR/SSIM metrics in Table 2. This experiment can show the effectiveness and interpretability of our approach.

	Decomposition	Simulation
bandpass#1: 1000 nm	35.83 / 0.947	50.09 / 0.984
bandpass#2: 1050 nm	45.27 / 0.983	52.88 / 0.997
bandpass#3: 1100 nm	57.34 / 0.996	48.68 / 0.982
bandpass#4: 1150 nm	45.89 / 0.807	50.51 / 0.983
bandpass#5: 1200 nm	65.11 / 0.997	51.89 / 0.981

Table 2. The table displays the PSNR/SSIM of the output image of the Decomposition-Net and Simulation-Net with different bandpass wavelengths.

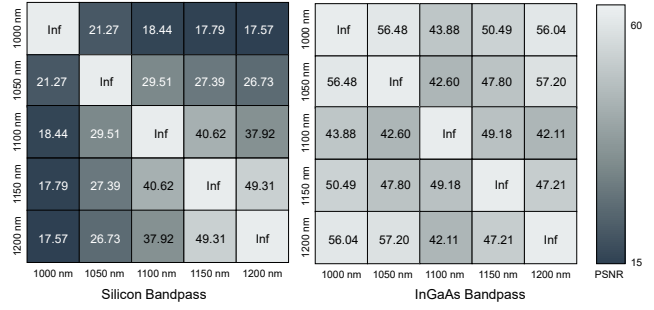


Figure 11. The left and right figure separately correspond to the Confusion Matrix of the pictures from our dataset imaging with Silicon sensor and InGaAs sensor intercepted by filters.

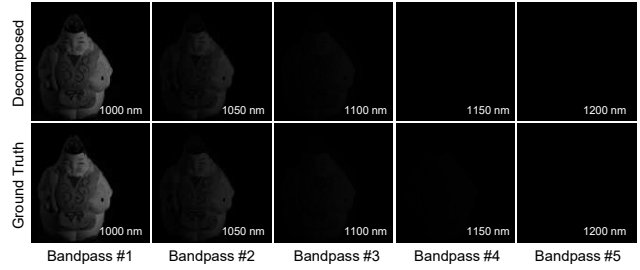


Figure 12. Experiment results of the Decom-Net for interpretability and effectiveness. The first row shows the outputs of the Decom-Net and the second row shows the ground truth which is intercepted by the real filters in each wavelength band.

## 7. Conclusion & Discussion

**Conclusion.** In this paper, we propose a novel solution which combines ingenious optical design and advanced image processing algorithms for SWIR imaging using a Silicon camera with a longpass filter. We design a multi-channel imaging system and build a new dataset called Silicon-to-InGaAs (SI). Using the new dataset, we design a proof-of-concept prototype using a multi-branch fully-convolution network. The experimental results demonstrate our proposed solution is effective and practical. As far as we know, this is the first attempt in SWIR imaging field.

**Benefits.** Our approach avoids the use of expensive SWIR sensors, therefore, significantly reduces hardware costs. And it retains the advantages of Silicon cameras such as high resolution, mature workmanship, well support for dynamic scenes and so on.

**Limitations.** Our method analyzes and simulates the shorter wavelength region ( $<1200nm$ ) of SWIR while approximating the rest, which may be solved with more advanced devices or optic design to prove our approach.

**Future work.** In the future, we will continue yielding improvements in imaging quality and optimizing runtime to promote the industrial process. Other future work aspects include dynamic scenarios optimization, specific applications (such as skin detection, surveillance) and so on.



## References

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016. 7
- [2] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathématique*, 346(9-10):589–592, 2008. 2
- [3] Huaijin Chen, M Salman Asif, Aswin C Sankaranarayanan, and Ashok Veeraraghavan. Fpa-cs: Focal plane array-based compressive imaging in short-wave infrared. *CVPR*, pages 2358–2366, 2015. 2, 3
- [4] François Chollet et al. Keras. <https://github.com/keras-team/keras>, 2015. 7
- [5] Weisheng Dong, Guangming Shi, and Xin Li. Nonlocal image restoration with bilateral variance estimation: A low-rank approach. *IEEE Trans. Image Processing*, 22(2):700–711, 2013. 3
- [6] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006. 1, 2
- [7] MP Doublier, A Roache, and S Potel. *Application of SWIR spectroscopy in very low-grade metamorphic environments: a comparison with XRD methods*. Geological Survey of Western Australia, 2010. 3
- [8] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008. 2, 3
- [9] H Figgemeier, M Benecke, K Hofmann, R Oelmaier, A Sieck, J Wendler, and J Ziegler. Swir detectors for night vision at aim. In *Infrared Technology and Applications XL*, volume 9070, page 907008. International Society for Optics and Photonics, 2014. 3
- [10] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Flexible voxels for motion-aware videography. In *European Conference on Computer Vision*, pages 100–114. Springer, 2010. 3
- [11] Marc P Hansen and Douglas S Malchow. Overview of swir detectors, cameras, and applications. In *Thermosense Xxx*, volume 6939, page 69390I. International Society for Optics and Photonics, 2008. 3
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5
- [13] Jason Holloway, Aswin C Sankaranarayanan, Ashok Veeraraghavan, and Salil Tambe. Flutter shutter video camera for compressive sensing of videos. In *Computational Photography (ICCP), 2012 IEEE International Conference on*, pages 1–9. IEEE, 2012. 3
- [14] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 5
- [15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017. 5, 8
- [16] Younghyun Jo, Seoung Wug Oh, Jaeyeon Kang, and Seon Joo Kim. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. *CVPR*, pages 3224–3232, 2018. 3
- [17] Jun Ke and Edmund Y Lam. Object reconstruction in block-based compressive imaging. *Optics express*, 20(20):22102–22117, 2012. 2
- [18] Ronan Kerviche, Nan Zhu, and Amit Ashok. Information-optimal scalable compressive imaging system. In *Computational Optical Sensing and Imaging*, pages CM2D–2. Optical Society of America, 2014. 2
- [19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7
- [20] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *CVPR*, 2(3):4, 2017. 3, 5
- [21] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J Brady. Coded aperture compressive temporal imaging. *Optics express*, 21(9):10526–10545, 2013. 3
- [22] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 5
- [23] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *IJ-CAI’81*, pages 674–679, 1981. 5
- [24] A Mahalanobis, R Shilling, R Murphy, and R Muiise. Recent results of medium wave infrared compressive sensing. *Applied optics*, 53(34):8060–8070, 2014. 2
- [25] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016. 5
- [26] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005. 2
- [27] Ying Qu, Hairong Qi, and Chiman Kwan. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2511–2520, 2018. 3
- [28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 5, 8
- [29] Aswin C Sankaranarayanan, Christoph Studer, and Richard G Baraniuk. Cs-muvi: Video compressive sensing for spatial-multiplexing cameras. In *Computational Photography (ICCP), 2012 IEEE International Conference on*, pages 1–10. IEEE, 2012. 3
- [30] Jian Wang, Mohit Gupta, and Aswin C Sankaranarayanan. Lisens-a scalable architecture for video compressive sensing. In *Computational Photography (ICCP), 2015 IEEE International Conference on*, pages 1–9. IEEE, 2015. 2
- [31] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6, 7, 8