QIRT'2022

10.21611/qirt.2022.2017

16th Quantitative InfraRed Thermography Conference

Simulation of single-pixel IR camera with CNN reconstruction algorithm

by Sebastian Urbaś*, Piotr Więcek**, Bogusław Więcek*

* *Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, Poland*
** *TEXO Systems, Telefoniczna 23G/29, 91-728 Łódź, Poland*

Abstract

The article describes an innovative concept of a single-pixel thermal camera based on compressing sensing and deep learning image reconstruction. The proposed approach uses a deep learning network to reconstruct an image from sparse data. The network architecture and preliminary results for various compression ratios are presented. Compressed data was obtained using random and deterministic spatial modulation of thermal images. In this research the modulation was realized from original thermal images by software means. The results presented in this paper were obtained for low-resolution 80x80 infrared images.

1. Introduction

Nowadays, modern thermal cameras mostly use an array of thousands of pixels to obtain high-resolution images. New promising, low-cost approach to capture thermal images is SPI (*Single-Pixel Imaging*) based on *Compressed Sensing* (CS) methodology [1]. An advantage of use CS algorithm is possibility to reconstruct of IR images in various spectrum ranges e.g. LWIR (*Long-Wave InfraRed*) 8-12 μm or MWIR (*Medium-Wave Infrared*) 3-5 μm. It depends on the spectral sensitivity of the thermal detector used in the IR camera [9]. Most of the Single Pixel Cameras (SPC) uses a sequence of random, Hadamard or Fourier mask patterns, to filter and compress the scene, which total intensity is integrated by a single detector [1], [6].

In this research *Convolutional Neural Network* (*CNN*) was implemented using *decoder* architecture for image decompressing. In his work, the spatial modulation of thermal images was simulated by a software modeling different shutters for thermal scene compression. We propose to use different shapes of the shutter for data compression in form of the moving *curtains*. This concept was already implemented and verified in *Matlab* environment using well-known in literature reconstruction algorithm - *L1 magic* [1], [6].

Thermovision allows the registration of thermal radiation emitted by physical bodies. Each body that has its own temperature higher than absolute zero ($0K \approx -273,15°C$) emits infrared radiation. The most important element of each thermal imaging camera is the infrared (IR) detector. The primary task of the IR detector is to convert infrared radiation into an electrical signal. There are currently two main groups of IR detectors: bolometric and photon detectors [5]. An important parameter of the infrared detector is its spectral range of absorption. Currently, most of infrared systems use two spectral ranges: LWIR (*Long-Wave InfraRed*) $8-12$ μm and MWIR (*Medium-Wave InfraRed*) $3-5$ μm [5], [9]. Thermal imaging cameras are expensive mainly due to the use of matrix infrared detectors and IR optics. Single-detector imaging is presently a new alternative to matrix-based IR systems. Currently, it is a solution mainly for laboratory applications.

The preliminary research presents the results of simulations of an IR camera with a single detector. The use of a single detector and dedicated spatial shutters in front is closely related to the measurement of sparse signals. In this article, the CNN with *decoder* architecture and the *L1-magic* algorithm were used to reconstruct IR images from sparse signal. The *L1-magic* algorithm is often used in the CS image reconstruction from incomplete measurements. As the sampling theorem states, the signal can be recovered back when the sampling frequency is greater than twice of the highest frequency component of the signal. It is not a case in Compressive Sensing measurements [1]. Using the CS method allows the signal or image to be reconstructed from much small number of measurements below the Nyquist criterion. In order to reconstruct sparse signal by using CS method it is possible to use different algorithms. In case of this research, the *L1-magic* algorithm was used, based on minimizing the l_1-norm [1].

The main purpose of the presented research was to use the CNN network with the *decoder* architecture to reconstruct infrared images from the sparse vector. The concept of CNN was created several decades ago [3], [4]. It assumes the use of an algorithm to train a specialized multilayer network. Currently, CNN learning systems are being used in image processing on a large scale. Nowadays, many varieties of networks have emerged, that are a modification of the basic structure of the CNN network [3].

2. Materials and methods

2.1 Single pixel IR camera

The figure 1 shows the principle of operation of the simulated SPC IR camera. The object emits infrared radiation. Before the IR radiation is focused on the single detector using lens, it passes through the *SLM (Spatial Light Modulator)*, which

may be implemented as the *curtain*-type shutter proposed by the authors. Originally, the random mask patterns shutter was mainly being used by the other researchers [1]. In result, the averaged radiation is measured by a single-detector, typically by a bolometer or an IR photodiode. The measurements stored as the compressed data vectors are then processed to reconstruct an IR image by the previously mentioned the *L1-magic* algorithm or by using CNN image reconstruction algorithm with the *decoder* architecture - figure 2.

In the target single-pixel IR camera, the part of the proposed algorithms is planned to be implemented in a hardware way using a *FPGA* chip. There are *FPGA* integrated circuits with several hundreds to even several millions logical modules and memory elements, as well as multiple-bit adders and accumulators. In order to develop the system, the hardware description languages such as *VHDL* or *Verilog* are used to configure the *FPGA* [7].

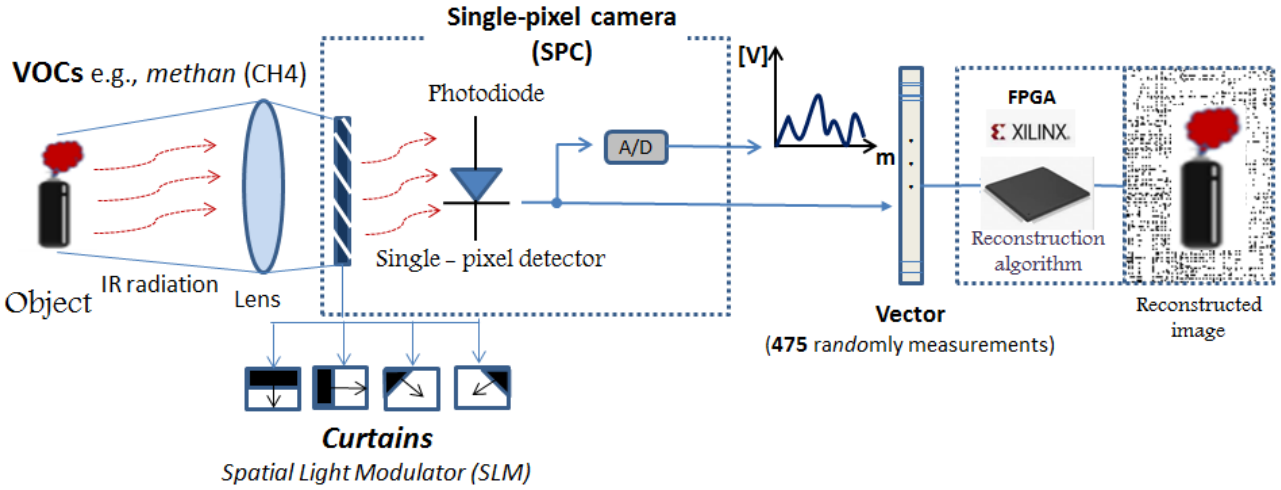

Fig. 1. *Block diagram of the single-pixel IR camera*

2.2 Convolutional neural network architecture

In this research, the *autoencoder* (AE) architecture was used - figure 2 [4]. The most important part of the *autoencoder* is the sparse data vector sometimes called *bottleneck*, which contains the compressed knowledge representations. The important part of the *autoencoder* is the *decoder*, which was used in the simulation carried out during this research. This is a module that helps the network to *"decompress"* the knowledge representations and reconstructs the data back from its original form.

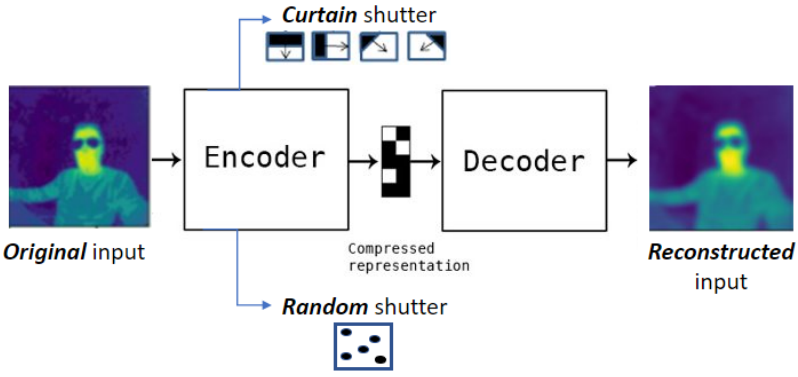

Fig. 2. *Autoencoder concept*

The *encoder* was simulated by using two different shutters - one using the set of 4 curtains (vertical, horizontal, diagonal and anti-diagonal) and the second shutter with many randomly distributed openings.

As a result of the use of these shutters, the compressed vectors were generated. Appropriately modified shutters will give a properly compressed vector with 475, 1000 or 2000 data samples. The use of the CNN *decoder* allows the image to be reconstructed from a compressed vector. The *decoder* structure was implemented using open-source software *Keras* in version 2.8.0 [10]. This is the library that provides effective Python interface for programing deep learning]artificial neural networks. The *Keras* toolbox works directly as an interface to the *Tensorflow* library and it has ready-made implementations of frequently used modules in the machine learning [13]. In addition, it has the necessary features that have been used in the implementation of the decoder. This library supports Convolutional Neural Network and it provides useful tools to implement e.g.: activation functions, optimizers, dropout, batch normalization, and pooling. In the presented architecture of the *decoder*, the *ReLU* activation function was used.

In order to estimate the results of the IR image reconstruction for each algorithm, it was assumed that there was no need to use a physical spatial modulator for data compression. The compression was performed by a software using thermal images. In order to obtain thermal images two IR cameras was used - 80x80 pixel system with a-Si detector *Micro80Gen2* [11] and 640x480 equipped with *VOx Imager* [12].

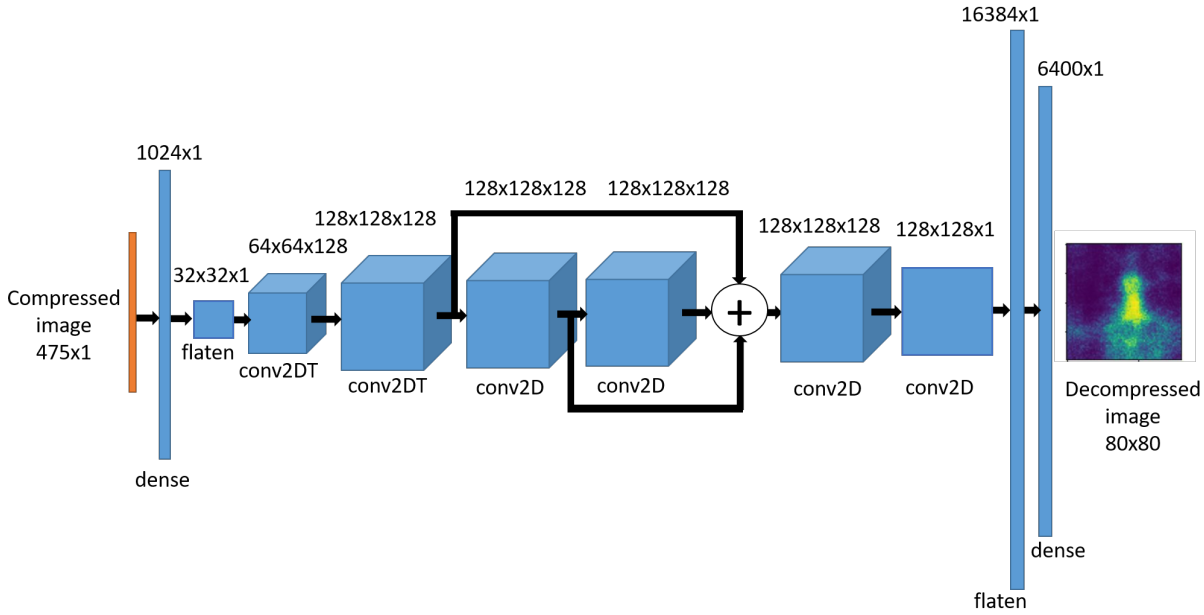

Fig. 3. *Architecture of Decoder CNN Network with residual connections*

The decoder architecture proposed in this research is shown in figure 3. It consists of $conv2DT$ ($transposed2Dconvolution$) and classical $conv2D$, flaten and dense layers, with local and global residual connections. Transposed 2D convolution layers ensure the features' up-sampling and helps final image generation.

Table 1. *The structure of the network*

Hidden layers No.	Type of the layer	Number of kernels
1	Dense Layer	None
2	Reshape/flaten	None
3	Conv2DTranspose Layer	128
4	Conv2DTranspose Layer	128
5	Convolutional Layer	128
6	Convolutional Layer	128
7	Convolutional Layer	128
8	Convolutional Layer	1
9	Flatten Layer	None
10	Dense Layer	None

Table 1 shows the types of layers used in the proposed CNN with their depth - the number of kernels (filters) used.

2.3 Acquisition of thermal images for training

In order to acquire thermal images to train convolutional neural networks, a low resolution thermal imaging camera $\mu IR80$ was used, made by the co-author of this article during diploma thesis [7]. This thermal imaging camera has the a-Si (amorphous silicon) microbolometer detector with resolution of 80x80 bolometers [11]. The typical bolometer detector is a square or rectangular matrix of bolometers connected internally to a readout circuit [11]. The IR camera is equipped with f=1.9 mm lens. The size of the detector is 34 μm. It ensures the *Instantaneous Field of View (IFOV)* of about 18 mrad.

The infrared detector *Micro80Gen* is made in a *BGA* housing. Single ball diameter is 0.5 mm and ball spacing is 1 mm. The detector has 107 soldering pins. The use of *BGA* housing in modern microbolometric detectors means that a large number of pins does not take up much space. The detector was soldered to the PCB using the *WEP 853AA* soldering station. The maximum temperature of the detector housing without fear of damaging is $\approx 260^{\circ}C$, according to the manufacturer's documentation [11]. This type of housing is also used in mobile phones, tablets and computers.

The size of the IR camera presented in figure 4 has the small size 25x39 mm. For this reason, such an infrared camera can be used e.g. in UAVs (*Unmanned Aerial Vehicles*).

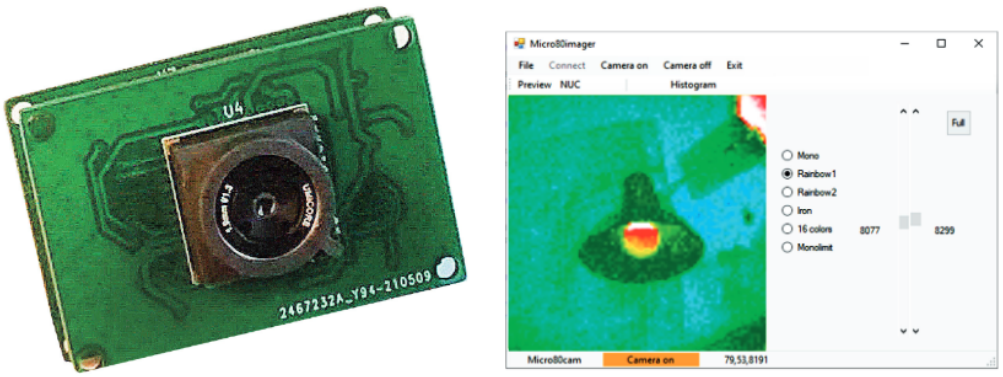

Fig. 4. *Developed 80 x 80 bolometric cameras μIR80 (left), screen of the Micro80imager application (rigth)*

Additionally, the high resolution, high-sensitivity thermal imaging camera *VOx Imager* was also used to generate thermal images with better quality and enlarge the training dataset. This thermal camera was equipped with high-sensitive 640x480 VOx microbolometer, manufactured by SCD *SemiConductor Devices* [12]. The camera sensitivity is less than 35 mK and the pixel pitch is equal 17μm. However, the training dataset uses thermal images with the low resolution of 80x80 pixels only. The C# application *Convert80* was written using *.NetFramework* platform to cut out subimages with 80x80 pixel resolution from 640x480 pixel images taken with *VOx Imager* as shown in figure 5. The dataset has been divided into the training and validation subsets with 1000 and 50 thermal images respectively.

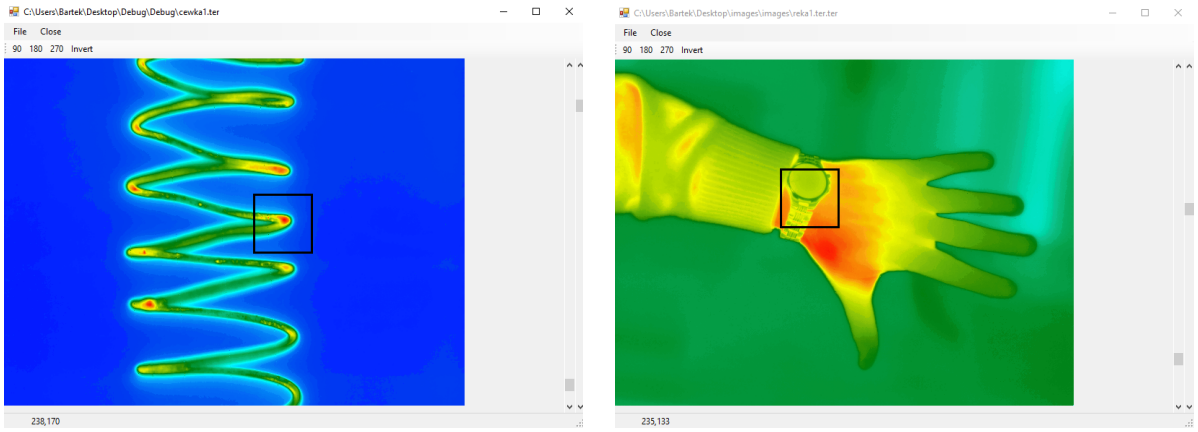

Fig. 5. *Application Convert80 for generating training dataset from an IR image with the high resolution of 640x480. The black box indicates the segmented area with the 80x80 pixels resolution*

3. Results and discussion

In order to assess objectively the decompressed IR images, the $PSNR$ (*Peak signal-to-noise ratio*) parameter was used to compare original and reconstructed thermal images. The $PSNR$ measure needs *Mean Squared Error - MSE* to be calculated first.

$$MSE = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} ([f(i,j) - f'(i,j)]^2) \tag{1}$$

where: *N, M* - thermal image dimensions in pixels, *f(i,j)* - pixel value with coordinates *(i, j)* of the original IR image, *f'(i,j)* - pixel value with coordinates *(i, j)* of the compressed IR image.

The $PSNR$ coefficient takes a form:

$$PSNR = 10 log_{10} \frac{L^2}{MSE} \tag{2}$$

where L is the range of pixel values.

In the case of both used IR systems: the $\mu IR80$ camera and the *VOx Imager*, each pixel is stored as 16-bit data. In consequence, $L = 65536$. Coefficient MSE is the mean square error between the reference and reconstructed thermal images. Figure 6 shows qualitatively the difference of the reconstructed images using both algorithms: *L1-magic* and *CNN-decoder*.

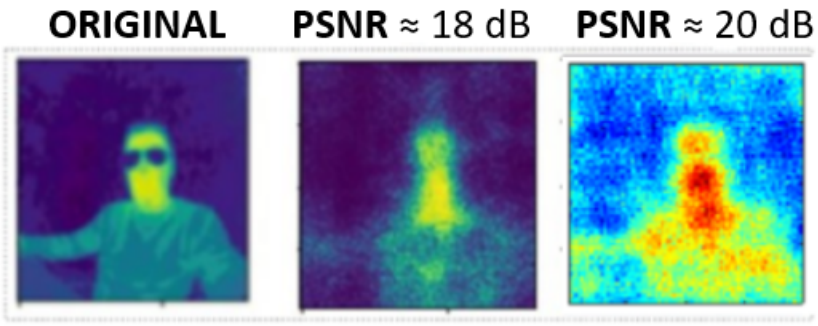

Fig. 6. *a) The original IR image taken by Micro80Gen thermal camera at 80x80 resolution.*
*b) The IR image reconstructed by the SPC using the CNN (**475** randomly selected measurements)*
*c) The IR image reconstructed with using L1 - magic algorithm (**475** randomly selected measurements)*

The compression ratio for both reconstruction algorithms is approximately 7,42 %. As one can see, this is high compression ratio but sufficient for reconstruction the original image. The reconstruction of the thermal images was performed using the processor (CPU) - *AMD EPYC 7301 16-Core Processor*, with clock frequency 2.20 GHz. The time to reconstruct the IR image using the *CNN-decoder* architecture is approximately equal to 2.3 s.

The table 2 shows the results of IR image reconstruction for different shutters using the *L1 magic* algorithm. The best result was achieved for the *Random2000* shutter. The $PSNR$ measure is equal to 35 dB. The compression ratio for this shutter is about

Table 2. *Averaged reconstruction results for 10 validation IR images for L1-magic algorithm*

Shutter type	Compression rate, %	PSNR, dB
Random475	7.42	23.173
Random1000	15,6	28.410
Random2000	31,2	35.081
Curtain475	7.42	21.105

Figure 7 presents the images reconstructed for different compression rates using classical *L1-magic* algorithm. The expected conclusion is that the lower the compression rate, the better reconstructed image. The reconstruction results of the same images for different compression rates using *decoder-architecture CNN* are shown in figure 8.

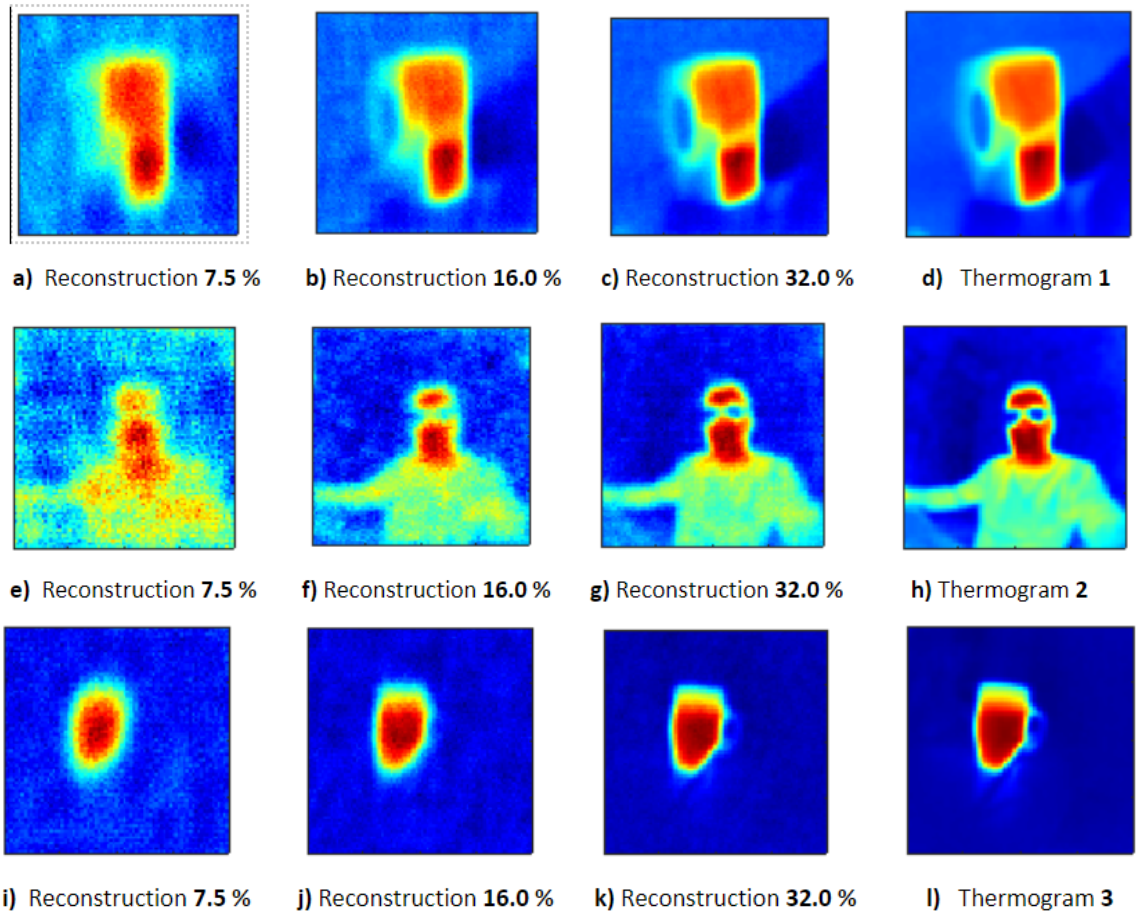

Fig. 7. *Results of reconstruction of a thermal images (80x80 pixels) normalized to (0 - 65535) scale for different compression rates using L1 - Magic algorithm , (a,e,i) 475-element compressed vector, (b,f,j) 1000-element compressed vector, (c,g,k) 2000-element compressed vector, (d,h,l) - original images*

The comparison of decompression results presented qualitatively in figures 7 and 8 as well as in tables 2 and 3 clearly confirms the superiority of the classical *L1-magic* approach over the CNN implementation of compressed thermal image reconstruction using SPC. The table 3 summarizes the results of IR images reconstruction for the same shutters with using *Convolutional Neural Network*. The obtained results of the reconstruction using CNN architecture are much worse than in the case of using the *L1 magic* algorithm. The best result od reconstruction for *Random2000* shutter is ~ 19 dB.

Table 3. *Averaged reconstruction results for 50 validation IR images and using CNN decoder architecture*

Shutter type	Epochs = 1000	Epochs = 2000	Epochs = 5000
Random475	16.973	17.699	18.683
Random1000	18.581	18.732	18.514
Random2000	18.391	18.542	18.721
Curtain475	14.964	14.990	15.203

In the opinion of the authors, the presented research results are preliminary showing and confirming the possibility of $CNNs$ applications for decompression problem of data generated by $SPCs$. Despite the poorer results obtained with the CNN method, the research seems promising and will certainly be continued.

According experience of the authors in the domain of artificial intelligence systems, the next step will lead to the significant data augmentation for CNN training. The presented results were obtained for 1000 images for training only and without any dataset augmentation. In addition, the transfer learning is planned in the future. The modification and

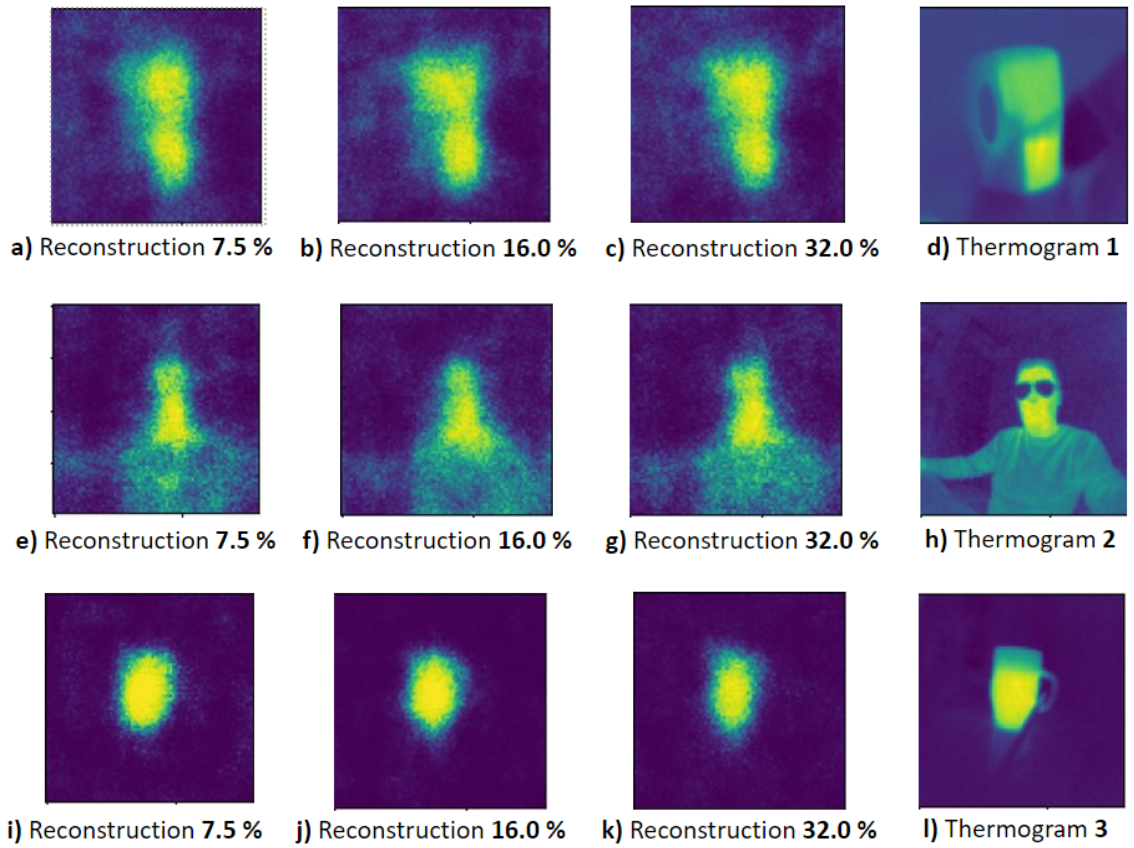

Fig. 8. *Results of reconstruction of a thermal images (80x80 pixels) on the normalized scale (0 - 65535) for different compression rates using CNN (a,e,i) 475-element compressed vector, (b,f,j) 1000-element compressed vector, (c,g,k) 2000-element compressed vector, (d,h,l) - original images*

optimization of CNN architecture will be considered as well.

4. Conclusions

The single pixel camera design makes it possible to build low-cost, low-power, small and high-quality imaging devices for a wide range of applications (e.g. remote imaging, hyperspectral imaging or video acquisition). The SPC requires a special optical setup consisting of a spatial light modulator and a single photodetector, which averages the incoming radiation. The compressed sensing theory provides perfect theoretical framework for single pixel imaging. It is based on the use of random SLM pattern and *L1-minimization* for image recovery. In order to reconstruct the infrared images, the autoencoder technique was used. *Autoencoders* always consist of two main components. The first is the *encoder* that transforms the input data into the compressed representation. In this research, the *encoder* was realized using a special software. In the next step of the research, the real *Spatial Light Modulator* will be developed to obtain a compressed vector. The second element is the *decoder*, the purpose of which is to reconstruct the IR images based on a compressed vector. In this research, the *Convolutional Neural Network* with the appropriate architecture plays the role of the decoder. The *Convolutional Neural Network* architecture was implemented using an open source machine learning framework - *Tensorflow*. The *L1 magic* algorithm was used for the purpose of comparative image reconstruction. After reviewing the results, it turns out that the *L1 algorithm* gives better results than image reconstruction using CNN. A physically constructed camera with a single detector can be applied to detect $VOCs$ (*Volatile Organic Compounds*) gases, mainly from hydrocarbon group e.g.(methan - CH_4) or carbon dioxide (CO_2).

References

[1] R. G. Baraniuk, "Compressive sensing," IEEE Signal Processing Mag., vol. 24, no. 4, pp. 118–120, 124, July 2007.

[2] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Prob. 23(3), 969–985 (2007).

[3] LeCun Y., Bengio Y., Convolutional network for images, speech, and time-series. 1995 in Arbib M. A. (editor), *The Handbook of Brain Theory and Neural Network*. MIT Press, Massachuesetts.

[4] Osowski S. *Deep neural networks in application to data mining*, Przegląd telekomunikacyjny, nr 5/2018, DOI: 10.15199/59.2018.5.2.

[5] Rogalski A., Infrared Detectors - second edition, CRC Press, 2020, ISBN 9780367577094.

[6] Szajewska A., Simulation of the Operation of a Single Pixel Camera with Compressive Sensing in the Long-Wave Infrared, „Pomiary Automatyka Robotyka", R. 25, Nr 2/2021, 53-60, DOI: 10.14313/PAR_240/53.

[7] Urbaś S., Więcek B., Development of Low-Resolution, Low-Power and Low-Cost Infrared System, „Pomiary Automatyka Robotyka", R. 25, Nr 2/2021, 47-52, DOI: 10.14313/PAR_240/47.

[8] Urbaś S., Development of infrared cameras (Projektowanie kamer termowizyjnych), M.Sc. diploma thesis, Institute of Electronics, Lodz University of Technology, academic year 2020/2021.

[9] Więcek B., Pacholski K., Olbrycht R., Strąkowski R., Kałuża M., Borecki M., Wittchen W., Termografia i spektrometria w podczerwieni : zastosowania przemysłowe, PWN, Warszawa, 2017, ISBN/ISSN: 9788301191870, (in Polish).

[10] *https://keras.io/*, (03.06.2022)

[11] $Micro80Gen2^{TM}$ - LynRed technical documentation, *https://lynred.com/*.

[12] *https://www.scd.co.il/wp-content/VOxImager/brochureILV7.pdf* (23.05.2022) - technical documentation [on-line]

[13] *https://www.tensorflow.org/*, (03.06.2022)