DCNN Optimization Using Multi-Resolution Image Fusion

Abdullah A. Alshehri[1], Adam Lutz[2], Soundararajan Ezekiel[2*], Larry Pearlstein[3], and John Conlen[2]

[1]Faculty of Engineering, King Abdulaziz University
Jeddah, 21589 KSA
[e-mail: eng.dean.rabigh@kau.edu.sa]

[2] Department of Mathematical and Computer Sciences, Indiana University of Pennsylvania
Indiana, Pennsylvania 15701, USA
[e-mail: sezekiel@iup.edu]

[3] Electrical and Computer Engineering Department, The College of New Jersey
Ewing Township, NJ 08628 USA
[e-mail: pearlstein@tcnj.edu]

[*]Corresponding author: Soundararajan Ezekiel

Received October 11, 2020; accepted November 3, 2020; published November 30, 2020

Abstract

In recent years, advancements in machine learning capabilities have allowed it to see widespread adoption for tasks such as object detection, image classification, and anomaly detection. However, despite their promise, a limitation lies in the fact that a network's performance quality is based on the data which it receives. A well-trained network will still have poor performance if the subsequent data supplied to it contains artifacts, out of focus regions, or other visual distortions. Under normal circumstances, images of the same scene captured from differing points of focus, angles, or modalities must be separately analysed by the network, despite possibly containing overlapping information such as in the case of images of the same scene captured from different angles, or irrelevant information such as images captured from infrared sensors which can capture thermal information well but not topographical details. This factor can potentially add significantly to the computational time and resources required to utilize the network without providing any additional benefit. In this study, we plan to explore using image fusion techniques to assemble multiple images of the same scene into a single image that retains the most salient key features of the individual source images while discarding overlapping or irrelevant data that does not provide any benefit to the network. Utilizing this image fusion step before inputting a dataset into the network, the number of images would be significantly reduced with the potential to improve the classification performance accuracy by enhancing images while discarding irrelevant and overlapping regions.

Keywords: Image Fusion, Deep Convolutional Neural Networks, Wavelets, Image Classification, Heterogeneous DCNN Fusion

1. Introduction

In recent years, machine learning has seen widespread adoption in fields such as classification, object recognition, and sentiment analysis. Advancements in the computational capabilities of GPUs has allowed machine learning techniques to leverage massively parallel processors, increasing practical viability to analyze and process data at rates far beyond those attainable through the use of a CPU alone. Specifically, Deep Convolutional Neural Networks (DCNNs) are able to perform high-level classification tasks involving text, images, and audio at rates that have been previously impossible [1,2]. This has allowed for the efficiency of classification to be increased such that some tedious, repetitive tasks can be significantly automated, thereby replacing or augmenting the error-prone human classification process. However, despite their promise, neural networks still suffer from several limitations, many of which relate to their need for an enormous quantity of representative training data [3]. Generally, the training process consumes an extreme amount of computational and storage resources, the subsequent performance of the network depends on the quality of the training data that is given. Especially in the case of visible light imagery, where out-of-focus sections, sensor malfunctions, and other distortions can significantly impact the raw pixel data given and the consequent integrity of the network. Furthermore, a substantial amount of training data is normally required for a well-trained image classification model. Images captured from a variety of angles, focuses, or different types of sensors are generally required, which can considerably add to the computational requirements and time necessary to test the network adequately. Normally, each individual image would have to be considered separately by the network, even if of the same scene but captured from a different angle or with another type of sensor.

One possible solution is by utilizing image fusion techniques to consolidate images of the same scene but captured from different focus, sensors, or modalities to diminish the resource requirements to deploy the network. By fusing images into a single stream of data to test a neural network model, images that have been captured from various angles, or with multi-modal or multi-spectral sensors can be analyzed simultaneously rather than being considered separately [4,5,6,7]. Multisensory applications were first created as a subset of remote sensing and have been developed as a subset of data fusion. Sources of fusion from video, audio, and numerical data is known as abstraction-wise data fusion. When applying this to cyberspace, this concept of data fusion profoundly relates to the study at hand. One early application of fusion was in the medical field where respiratory and electrooculography signals were fused with electroencephalography signals to develop numerous fatigue models for patients. Other types of fusion techniques such as data-drive, pyramid-based, and wavelet based fusion methods became viable as further advancements progressed [8]. The primary techniques used in this study are multi-focus and multi-modal fusion which allows separate images of the same scene which contain distinct key features to be fused into a single image. The applicability for image fusion techniques has proliferated into fields such as medical image analysis, environmental monitoring, and remote sensing [9] by allowing images of the same scene captured from different sensors or modalities to be combined. A key aspect of this study is optimizing the identification of features and segmenting regions of interest for images captured from different focuses or modalities, which can be substantially more computationally complex [10].

Massive amounts of digital data are produced by news sources, multimedia user content, and mobile collections in the form of videos and images. This data is not typically categorized,

indexed, or stored in a manner convenient for image fusion. Archives, such as ones run by newspapers, museums, and libraries require teams of people to annotate and classify such images. Developing an algorithm for automated image fusion would significantly reduce the time required for downstream image analysis, classification, and tagging [37]. To simplify the process, edge detection can be used to determine objects within the images [38]. Image fusion is a continually growing field in image processing within the information fusion field that shows promise for many applications [39]. Image fusion uses image processing techniques to integrate complementary data and increase the amount of information contained in the image [40, 41]. Applications of image fusion include video tracking, decision support, situation awareness, target recognition, compensation for obscuration, simultaneous tracking and identification, dynamic scene analysis, and sensor design [42, 43, 44, 45, 46, 47, 48].

Many techniques are available for image fusion. Bin and Chao [49] proposed a technique that uses image fusion to combine discrete multiwavelet transformations. Piella, et al., [50] proposed a multiresolution fusion algorithm that allows for the imposition of data dependent consistency constraints by combined region and pixel level image fusion. In any fusion assessment, difficulty arises from the fact that there is no clear ground-truth on which to base the assessment.

Although numerous image fusion algorithms have been developed, there has been no clear distinguishing criteria established to determine which is best [52, 53]. To determine the quality of the image fusion, methods involving information theory, image features, structural similarity, and human perception have been used. Additionally, most fusion algorithms are optimized to fuse multiple images. One of the main difficulties in quantifying the effectiveness of an image fusion algorithm is that some techniques are more specialized for one application, causing them to be less effective in others. The relationship is still ambiguous between an adapted fusion algorithm, the input images, and the quality of the images. As such, more work is needed to determine why a certain algorithm works better in a given situation than others.
Multiple images of the same scene captured from varying focus points create a wide variation of contextual angles of the scene can be fused through multi-focus image fusion into a single stream of data. Different perspectives, such as background and foreground focus are able to be consolidated into a single fused image that retains the most significant features of both the fore- and backgrounds. An issue, however, lies in a lack of suitable datasets that not only contain a foreground and background focus image that can be classified, that also contain an in-focus image to act as a ground truth. One possible solution to this is through the use of synthetic data generation, in which renders can be generated in a controllable environment. When utilizing real-world datasets, often times it is difficult to capture images under ideal circumstances or with a wide range of conditions. Having a controlled, simulated environment allows for unique, outlier conditions to be simulated to avoid the need to place physical sensors in uncommon environments or capture chance events. In this study, we utilize Panda3D to create and render simulated 3D environments. Models that act as objects to be classified are placed within the simulated environment in random positions and orientations, which allows a wide variety of images to be captured. The objects were placed in both the foreground and background to ensure that the objects would be out of focus in both the foreground and background of the images. Additionally, because the simulation is 3D, one of the most significant advantages is that it allows a depth map to be generated in order to capture the depth of the environment and the objects in it, something that would be difficult if using datasets that only contain foreground focus and background focus images. Using this depth

map, a depth of field simulation is applied to the generated images, which allows for the focal point to be changed based on depth. This allows for an arbitrary number of train and test images to be created, while containing objects that the network is able to classify. Moreover, combining visible light imagery with other modalities and sensors, such as infrared imaging, of the same scene encompasses multi-modal image fusion. This allows features that are undetectable or less pronounced by infrared sensors due to their limitations that are otherwise excellent at capturing information undetectable by visible light imagery, to be fused with environmental details such as topographical features and abiotic elements. Reducing the number of images required to properly test a network while retaining the most significant information through the creation of fused imagery allows network to be utilized efficiently and minimizing the impact of noise and other distortions. This process involves combining images that contain different significant features into a single stream of data, which can potentially diminish problems that arise from data captured in substandard conditions or with malfunctioning sensors. The primary objective of this study is the creation of an efficiently optimized network that can actively account for unanticipated environmental circumstances that could affect the integrity of individual images. By condensing the total amount of images required to test the network while extracting and retaining the most salient features in the image while discarding irrelevant noise, a major benefit would be decreasing the computational requirements to test the network by reducing the number of images, while also removing errors that could be caused by noise or artifacts.

The remainder of this paper is organized as follows: Section II describes the technical background of the techniques used in this study including multi-resolution transformations, the machine learning techniques, as well as the fusion methods and types of images being fused. Section III summarizes the methodology of heterogeneous DCNN fusion as well as the multi-resolution image fusion techniques used in this study. Section IV shows the results of our study and visualizes the numerical data of the classification results for the fused imagery. Section V discusses the impact of our study including the viability and effectiveness while exploring the future direction of our work.

2. Technical Background

Popular scenarios for image fusion include multi-modal, multi-resolution, and multi-focal. Multi-modal analysis fuses images that were captured from different sensors, creating an image from the most salient features from each source. Multi-resolution analysis allows for the creation of a series of approximations of an image. Multi-focal analysis fuses images of the same scene with different focal points, creating an all-in focus image. These techniques will allow for the retention of the most significant features and then discovering the separation of blur, edge, and noise coefficients and then apply our deconvolution kernel to them, rather than the pixel values. It is useful to note that blur is most apparent to the human eye in the edges contained within the images. The multi-resolution analysis will mitigate the problem of pixel intensity values becoming distorted by the deblurring process. Next, frequency domain fusion and wavelets will be discussed.

2.1 Frequency Domain Fusion

Fusing visible light imagery, such as multi-focus images, or multi-modal imagery can be accomplished through frequency domain fusion [11]. This technique involves decomposing

input images into their multi-resolution coefficients using a discrete transformation. These decomposed coefficients are segmented and manipulated using the most optimal fusion techniques for the images and synthesized into a single image through an inverse transformation [12,13]. The decomposed coefficients can additionally have thresholding and other denoising techniques applied to enhance the images before fusion and reconstruction.

2.2 Wavelet

Wavelets are defined as a finite oscillation which has an average value of zero. The amplitude of this begins at a value of zero and oscillates a finite number of times and ends with a value of zero as well. For a function $\psi(x)$ to be classified as a wavelet, the two following equations must be held:

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{1}$$

$$\int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega = C_{\psi} \tag{2}$$

with $\psi(\omega)$ being the Fourier transform of the chosen wavelet function and C_{ψ} being the *admissible constant.* Mostly derived by Daubechies, there has been several wavelets constructed. Wavelets are categorized from discretely over a grid to continuously over time or space as well as being real or complex valued. The rescale and translation parameters are the two fundamental characteristics of wavelets. Given a base wavelet $\psi(\omega)$ the family of wavelets, $\psi_{j,k}(x)$ can be defined as:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{|j|}} \psi\left(\frac{x-k}{j}\right) \tag{3}$$

where j is the scaling variable and k is the translation variable. These characteristics are what allow wavelets to be used to detect abrupt changes in signals, making them suitable transforms for point-wise edge detection. Continuous wavelet transforms (CWT) are defined as the inner product of a wavelet $\psi(x)$ and a function $f(x) \in L_2(R)$, expressed as:

$$f, \psi_{j,k} = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{|j|}} \psi\left(\frac{x-k}{j}\right) dx \tag{4}$$

with the function f for the purposes of signal or image processing representing a signal or image that has had the wavelet transformation applied to it. Given that images and signals for machine learning purposes would be sampled as discrete-space functions rather than processed as continuous-space functions, the discrete wavelet transform (DWT) is in general used to process them. The DWT, similar to the CWT, of a function f denoted as G_{Ψ} is expressed as:

$$G_{\psi}(f, \psi) = \frac{1}{\sqrt{M}} \sum_{i,j=1}^{n} f(x)\psi_{j,k}(x) \tag{5}$$

with M being a scaling weight. The transform decomposes signals into their coefficients which capture frequency and spatial information. There are, however, several disadvantages that result from moving a continuous transform to a discrete one. Namely, the wavelet transform loses directionality and shift-invariance, making it able to detect edges and abrupt changes but is unable to detect contours as a single section in the case of images for instance. This lack of shift-invariance is due to the DWT's inability to transform shifted versions of f in the time domain to shifted versions of G_{Ψ} in the wavelet domain. For the purposes of this study, images that have been captured for the purpose of frequency domain fusion can be decomposed using various wavelet and other multi-resolution transforms such as contourlet, curvelet, or bandelet. After the images have been decomposed on multiple levels, these coefficients are fused through various methods such as maximum, minimum, mean, or principal component analysis. The wavelets applied in this study include Coiflet, biorthogonal, Meyer, Daubechies, Symlet, and Gaussian derivatives [14,15,16,17,18,19].

2.3 Multi-Modal Fusion

Multi-modal image fusion allows for images captured from different sensors, such as visible light and infrared thermal imaging, to be fused into a single image that retains the most salient features of the individual images [20]. However, due to these differing modalities, errors can potentially appear such as color artifacts or other distortions and as such multi-modal image fusion must be precisely implemented. Additionally, erroneous distortions can also appear from the fusion of topographical and morphological details as infrared sensors are not as capable of capturing these details, which could create artificial shadowing and other distortions [21].

2.4 Multi-Resolution Fusion

It is possible to create an approximation of a set of images through a technique known as multi-resolution analysis (MRA), which can then be fused [22]. Images are decomposed into their detail coefficients across multiple levels of resolution which allow the most salient features of the image to be isolated and extracted while reducing the effects of noise and blur by manipulating the edge coefficients of images as opposed to raw pixel data. By decomposing images into multiple levels of resolution, the effects of blurring and artifacts can be diminished [23]. MRA is defined as a sequence of closed subspaces, $V_n, n \in ¢$ in ψ in a containment hierarchy:

$$.... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset .... \tag{6}$$

The nested spaces contain an intersection with the zero function and a union that is dense in $L(¡)$,

$$\cap_n V_j = \{0\}, \overline{\cup_j V_j} = L^2(R) \tag{7}$$

The hierarchy (8) is constructed such that *V*-spaces are self-similar,

$$f(2^j x) \in V_j \text{ if } f(x) \in V_0 \tag{8}$$

and there is a scaling function $\psi \in V_0$ whose integer-translates span the space V_0,

$$V_0 = \left\{ f \in \mathrm{L}^2(\mathrm{R}) \mid f(x) = \sum_k c_k \psi(x-k) \right\} \tag{9}$$

and for which the set $\{\psi\}$ is an orthonormal basis. By decomposing multiple images of the same scene rather than manipulation of raw pixels, the detail coefficients which contain the key features of the image can be manipulated and fused. This allows MRA-based image fusion to be utilized for both denoising and deblurring as well as image fusion.

2.5 Multi-Focus Fusion

Images of the same scene captured from differing points of focus can result in significant variations of the visual quality of different regions of the image. The capability of cameras to focus on key features without a loss of quality in other sections remains one of the most substantial constraints in image processing. A possible solution to this issue is using multi-focus image fusion, in which multiple images containing different points of focus are fused into a single, in-focus image. An image can have a region in-focus in various aspects of the frame, such as the background, foreground, mid-section, etc. while the remaining sections are out of focus, essentially containing useless information. By fusing the most prominent features of the in-focus regions, the resulting image would become subsequently have a better quality than the source images. Tenenbaum gradient (Tenengrad), spatial frequency (SF), sum-modified-Laplacian (SML) are types of focus-measurements that can be used to measure the clarity of regions of an image [24]. Additionally, information fusion applications are supported by numerous image fusion techniques [25]. Multi-focus image fusion has viability applicability for a wide range of fields, including aerial reconnaissance and surveillance, three-dimensional reconstruction, photography, and video production [26].

Fig. 1. Foreground focus (left), background focus (center), and fused (right)

As exampled in **Fig. 1**, multi-perspective fusion is the process of fusing multiple images of the same scene or object captured from different focal points.

2.5 Multi-Focus Fusion

One of the fusion methods utilized in this study is achieved by selecting the minimum criterion using the absolute value of the matrices that represent the source images, defined as:

$$f_{ij} = \begin{cases} a_{ij} & if \left|a_{ij}\right| \le \left|b_{ij}\right| \\ b_{ij} & otherwise \end{cases} \tag{10}$$

Due to it being possible for the multi-resolution coefficients in the matrix to be negative, the absolute value must be used as the pixel values must be positive. The purpose of this function is to compare the two entries of matrices a and b and select the lower value of the two for the corresponding matrix f .

2.5 Multi-Focus Fusion

Paired with the selection of the minimum it is also necessary to select the maximum. Like the minimum equation, the two input matrices that represent the source images are compared with the higher pixel value being selected to be used in the fused matrix f , defined as:

$$f_{ij} = \begin{cases} a_{ij} & if \left|a_{ij}\right| \ge \left|b_{ij}\right| \\ b_{ij} & otherwise \end{cases} \tag{11}$$

2.5 Multi-Focus Fusion

Another technique utilized for image fusion is using principal component analysis (PCA), which is a multi-variate analysis technique that is typically used for the dimensionality reduction of large matrices and feature extraction [27,28,29]. This process involves reducing large matrices of correlated variables into their corresponding, uncorrelated principal components. These uncorrelated, linearly independent components are ordered according to their variance and contain the most significant features of the data. As the variables are linearly independent, PCA is an orthogonal linear transformation that transforms a matrix X with n rows, that represent the source image, into a set of vectors of weights w with m columns that correspond to the input image, defined as:

$$w_{(k)} = (w_1, w_2, ...w_m) \tag{12}$$

The vectors of weights contain the principal component scores t mapped from each row of X where t is defined as:

$$t_{(1)} = (t_1, t_2, ..., t_l) \tag{13}$$

where $t_{k(i)} = x_{(i)} \cdot w_{(k)}$ for $i = 1,...,n \;\; k = 1,...,m$. The principal components are ordered in such a way that the first principal component contains the highest variance and thus represents most of the data, with the remaining k^{th} PCs ordered from highest to lowest variance. The first principal component, $w_{(1)}$ maximizes the variance by satisfying the following condition:

$$w_{(1)} = \underset{\|w\|=1}{\operatorname{argmax}}\{\|Xw\|^2\} = \underset{\|w\|=1}{\operatorname{argmax}}\{w^T X^T Xw\} \tag{14}$$

With the succeeding k^{th} PCs being found by subtracting the first $k-1$ principal components from the matrix x such that:

$$\hat{X}_k = X - \sum_{i}^{k-1} Xw_i w_i^T \tag{15}$$

followed by calculating the weight vector that maximizes the variance:

$$w_{(k)} = \underset{\|w\|=1}{\operatorname{argmax}}\{\left\|\hat{X}_k w\right\|^2\} = \underset{\|w\|=1}{\operatorname{argmax}}\{\frac{w^T \hat{X}_k^T \hat{X}_k w}{w^T w}\} \tag{16}$$

Every pair of principal components are orthogonal to each other as each are derived from the eigenvectors of the covariance of the data which are always symmetric, making PCA an orthogonal linear transform [30].

2.5 Multi-Focus Fusion

Deep Convolutional Neural Networks (DCNNs) are a prominent type of learning model normally composed of a deep, feed-forward architecture that can learn the features of their input. Based on the features learned from the training data, a well-trained model can classify new input into the predefined labels. Because new input must be classified into the categories or labels that the model was trained on, these networks are specific to the data on which they are trained [31]. The feed-forward architecture of DCNNs allows the output of previous layers to be used as input for subsequent connected layers, in this regard the architecture is a variable number of layers stacked on one another [32], shown in **Fig. 2**. All parameters, such as the filters and weights of the network, aside from static variables such as layers or kernel size are initialized to random values. The network then receives input data and goes through a convolutional layer in which the network convolves over the input data to learn its features. An activation map which contains the output of each convolution over the entire input is calculated from the filters that convolve over the data [34]. The filters in the convolution layer are used to learn its input by activating when specific features correlated to classes are activated. The output of this convolutional step is then forward propagated through the rest of the network layers, using the output of previous layers as input for the next. The locations of features that were detected from the images are then mapped in the pooling layers [35]. The activation mappings which include the learned features of the network from the previous layers are then fed to the fully connected layers which are responsible for the high-level reasoning and classification of the network using the learned weights. A summation of the error is calculated across all the classes at the output layer and using gradient descent, the filter values and weights are updated through a backpropagation step. This backpropagation step allows the model to optimize its parameters to improve its classification accuracy.

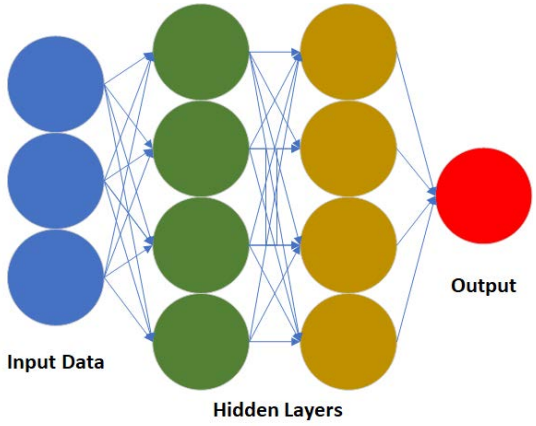


Fig. 2. Neural Network Architecture

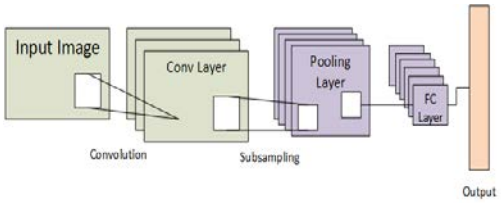


Fig. 3. Topology of a DCNN

2.5 FC_7 Layer Extraction

Prior to the softmax output layers of the neural networks utilized in this study, which provide the probabilities for each of classes of the input, are the fully connected layers which handle the high-level reasoning and correlations. The fully connected layers receive the output of all the previous layers, the convolutional and pooling layers that learn and map the features of the input, as the activation maps of the high-level features. The fully connected layers themselves are comprised of vectors that represent the probabilities of each label by determining how classes correlate to each of the features. The neural networks utilized in this study, despite having different architectures, share the property of having the FC_7 layer as their penultimate layer, shown in **Fig. 3**, which contains the activation weights across all the classes of each of the features. The properties of this layer allow it to be extracted before classification and used as input for a Support Vector Machine (SVM), as it contains the high-level correlations between features and classes.

2.5 Multi-Focus Fusion

Although differing networks having varying image classification architectures, a common property is the presence of fully connected layers of the same size. These layers receive the output of the previous layer into a fusible layer of size 4096. Heterogeneous DCNN Fusion takes advantage of the identical FC_7 layer sizes of the different architectures used by extracting these feature vectors and fusing them into a single feature vector that high-level reasoning and correlations of all the individual networks [36]. The extracted feature vectors can be fused using a variety of fusion methods, such as the maximum or minimum values, summation, or concatenation. This consolidates them into a single feature vector which can then be used as input for an SVM classifier, which is then used for classification, shown in **Fig. 4**.

Fig. 4. Heterogeneous DCNN Fusion

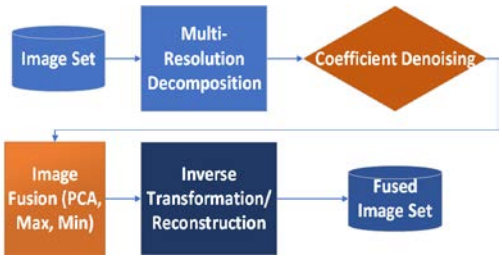


Fig. 5. Image Fusion Methodology

3. Methodology

For our study, multi-focus images were used, images containing multiple focal points were first generated, with the general image fusion methodology shown in **Fig. 5**. The images used in this study were synthetically generated and blurred between the foregrounds and backgrounds, creating two out of focus images from a single ground truth image, shown in **Fig. 6-8** which include the ground truth images, the out of focus images, and the reconstructed fused images. In order to gather a suitable amount of images that have multiple focal points, images were synthetic data generation was used to create 3D renders of two types of objects in the same environment in order to test binary classification accuracy. Two sets of images were generated, the actual 3D render as well as a generated depth map that was used to determine the foreground and background areas of the image. The objects were places randomly within the scene, both in the foreground and background, as well as rotated in order to obtain a variety of images. After the image sets were generated, the depth map was used to simulate different depth of fields using a moving average based on the maximum and minimum pixel values of the depth map. From this, two images, a foreground and background focus image were generated while still preserving the original images to act as ground truth. The images were then decomposed using a multi-resolution transform into their detail coefficients and a thresholding step applied for the purpose of denoising and image enhancement before fusing them, shown in Algorithm 1.

Algorithm 1 Wavelet Denoising

procedure WAVELETDENOISING

n = # of Images

M = Threshold Technique

for i : 1 → n

for each image I

[C, S] ← wavedec2*(*I_i*, level, waveletName)*

[approximateCoefficients, detailedCoefficients] ← separation*(C)*

d ← thresholding*(detailedCoefficients, Threshold Value, M)*

C' ← concatenate*(approximateCoefficients, d)*

denoisedImage ← waverec2(C*', S, level, waveletName)*

end

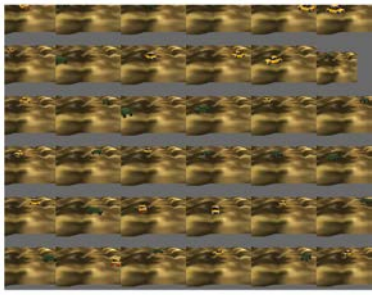


Fig. 6. Generated Ground Truth

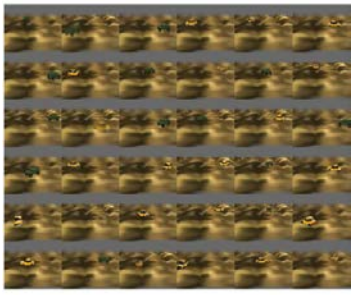


Fig. 7. Multi-Focus Images

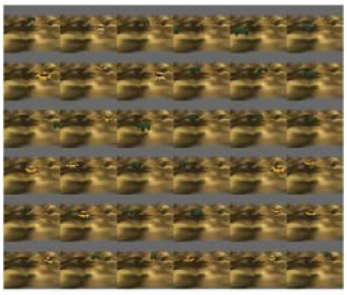


Fig. 8. Reconstructed Fused Images

The decomposed details coefficients were then fused to combine the most salient and significant features of each of the source images while reducing the amount of irrelevant information, caused by out of focus regions, or topographical information that thermal imaging is not adept at capturing. The decomposed images contain the directional coefficients which capture edges and directionality which make them ideal for detecting and fusing the most prominent features of the original images. For example, the in-focus regions of multi-focus images were typically found to have stronger detail coefficients compared to the out of focus regions, as seen in **Fig. 6**. Various fusion methods were applied to the denoised coefficients, including max, min, and PCA, to fuse them into a single stream of data that retains the most significant key features of the source images, shown in Algorithm 2. An inverse transformation was then applied to the fused detail coefficients to reconstruct the final fused image, which can then be used as input for the three neural networks utilized in this study, AlexNet, VGG16, and VGG19. The general methodology for the classification and DCNN fusion is shown in **Fig. 7**, in which the images were used as input for each of the individual neural networks, with their respective FC_7 layers being extracted before classification. The layers were then fused using the various fusion methods applied for heterogeneous DCNN fusion, fusing each of the three individual feature vectors into a single, fused stream of data that encompasses the high-

Algorithm 2 Image Fusion

procedure IMAGEFUSION

MM = Multi-Modal Image

MF = Multi-Focus Image

M = Threshold Technique

for $i : 1 \rightarrow n$

for each MM, MF

$[C_{1,2}, S_{1,2}] \leftarrow$ wavedec2*(*$[MM_i, MF_i]$*, level, waveletName)*

[approximateCoefficients$_{1,2}$*, detailedCoefficients*$_{1,2}$*]* ← separation*(*$C_{1,2}$*)*

$d_{1,2}$' ← denoise*(detailedCoefficients*$_{1,2}$*, Threshold Value, M)*

new d ← fusion*(*d_1', d_2', *Max, Min, PCA)*

new c ← concatenate*(a, new d)*

fusedImage ← waverec2*(new c, S, level, waveletName)*

end

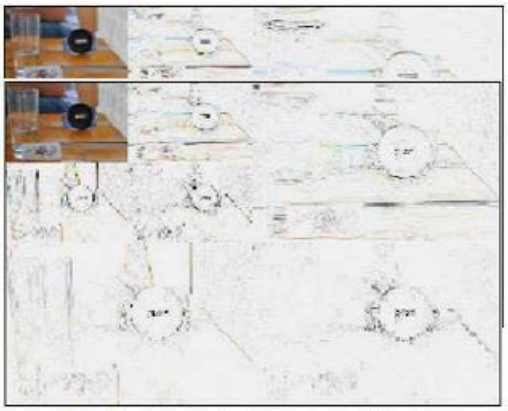


Fig. 9. Stronger background coefficients (top) & foreground coefficients (bottom)

Algorithm 3 DCNN Fusion

procedure DCNNFUSION(*reconstructedImageSet*)

n = # of layers

FC_{n-1} *= penultimate layer*

featureVectors ← FC_{n-1} Extraction(*reconstructedImageSet, AN,VGG16,VGG19*)

fusedFC$_{n-1}$ ← FEATUREFUSION(*featureVectors*)

classification ← SupportVectorMachine(*fusedFC*$_{n-1}$*,testImageSet)*

level reasoning and correlations of the source vectors, shown in Algorithm 3. The fused feature vector was then input into a SVM classifier which did the classification.

Algorithm 4

procedure SYNTHETIC DATA AUGMENTATION(dataset)

groundTruth ← (originalData)

fusionSet ←TrainFusion(originalImages, initParams)

blurredSet ←TrainBlur(originalImages, hiddenSize, regularzation)

combinedSet ← fusionSet + blurredSet

dataSets ← (original, blurredSet, fusionSet, combinedSet)

net1 = AlexNet; net2 = VGG16; net3 = VGG19;

[trainSet, testSet] ← Split(datasets{i}, groundTruth)

for i = 1:len(datasets) do:

procedure TrainFeatExtraction(trainSet, net1, net2, net3)

trainFC$_{n-1}$(A) ← ExtractFeatures(net1, trainSet)

trainFC$_{n-1}$(VGG16) ← ExtractFeatures(net2, trainSet)

trainFC$_{n-1}$(VGG19) ← ExtractFeatures(net3, trainSet)

procedure TestFeatExtraction(testSet, net1, net2, net3)

testFC$_{n-1}$(A) ← ExtractFeatures(net1, trainSet)

testFC$_{n-1}$(VGG16) ← ExtractFeatures(net2, trainSet)

testFC$_{n-1}$(VGG19) ← ExtractFeatures(net3, trainSet)

fusedTrainFeatures ← Max, Min, Sum, Mean(trainFC$_{n-1}$)

fusedTestFeatures ← Max, Min, Sum, Mean(testFC$_{n-1}$)

reducedTrainFeatures ← PCA, T-SNE(fusedTrainFeatures)

reducedTestFeatures ← PCA, T-SNE(fusedTestFeatures)

procedure ClassifyImages(reducedTrainFeatures, reducedTestFeatures)

classifier ← reducedTrainFeatures, labels

predY ← Predict(classifier, testFC$_{n-1}$(A))

acc = Mean(predY, testSet.lables)

predY ← Predict(classifier, testFC$_{n-1}$(VGG16))

acc = Mean(predY, testSet.lables)

predY ← Predict(classifier, testFC$_{n-1}$(VGG19))

acc = Mean(predY, testSet.lables)

predY ← Predict(classifier, reducedTestFeatures)

end

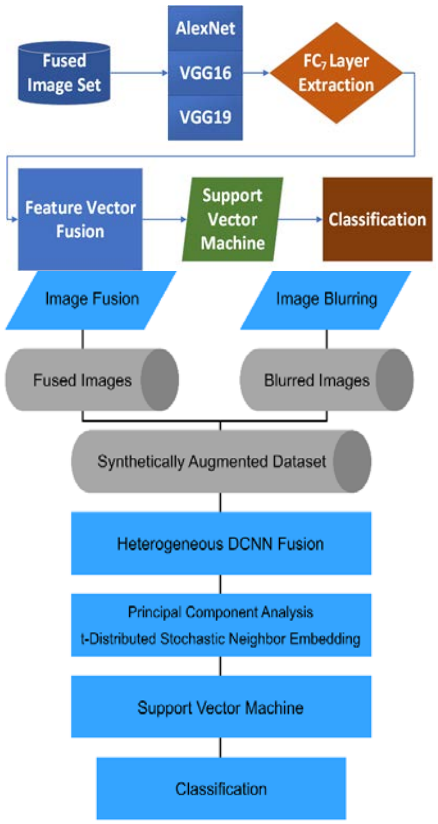


Fig. 10. DCNN Fusion & Classification Methodology

4. Results

Both multi-focus images that contained background and foreground focus images were utilized in our tests. Our trials attempted to compare the performance accuracy as well as the computational time of using non-fused images versus fused images of the same scene. Both the classification accuracy of the individual networks as well as the SVM trained using the heterogeneously fused feature vectors were found. The fusion methods that were used to fuse the feature vectors included maximum, minimum, average, and summation which all yielded uniformly sized feature vectors that could be input into an SVM classifier. Both accuracy and computational time were used as performance metrics to compare the two techniques. The results of the trials found that the fused image set had either the same or slightly higher classification accuracy for both the individual networks and the fused feature vector, as well as lower computational time, with the results shown in **Fig. 9-14** which show the accuracy of the trials using both fused and non-fused images. Our trials found that utilizing image fusion to fuse images into a single stream of data not only generally had either the same or better performance than the non-fused images, but also consistently had lower computational times to fully process the dataset, shown in **Fig. 16 & 17**, which compares the trial times of the fused image sets and unfused image sets.

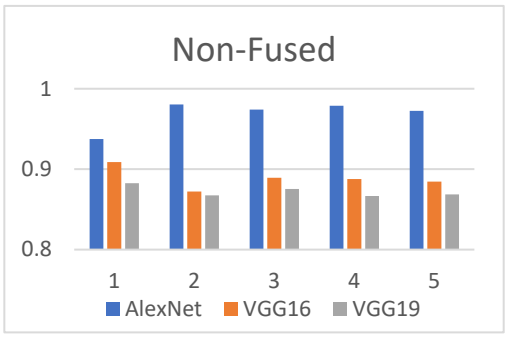


Fig. 12. Non-Fused Accuracy Results (Individual Networks)

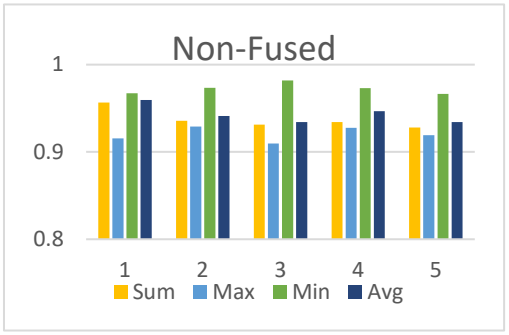


Fig. 13. Non-Fused Accuracy Results (DCNN Fusion)

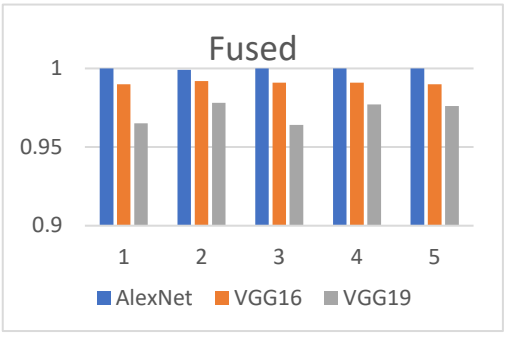


Fig. 14. Fused Accuracy Results (Individual Networks)

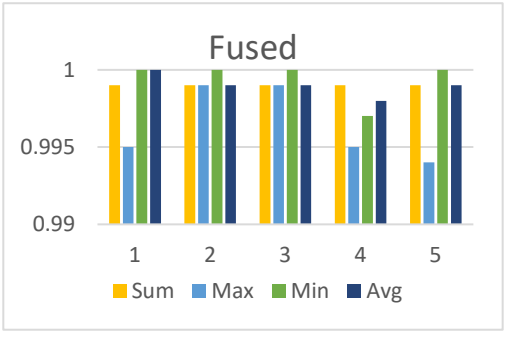


Fig. 15. Fused Accuracy Results (DCNN Fusion)

Our results found that compared to traditional methods, utilizing image as well as heterogeneous DCNN fusion yielded the same if not better accuracy, shown in **Tables 1 & 2** while reducing the total time required, shown in **Table 3**.

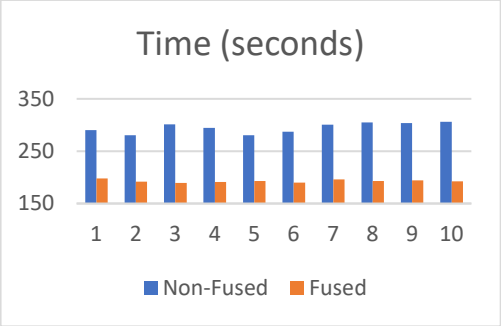


Fig. 16. Trial Time Results

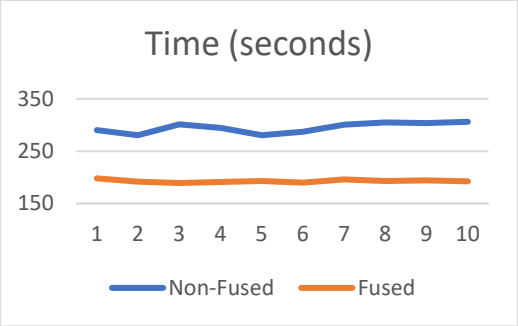


Fig. 17. Non-Fused Accuracy Results

Table 1. Non-Fused Accuracy Results (DCNN)

AlexNet	VGG16	VGG19	Sum	Max	Min	Avg
0.93762	0.90868	0.8827	0.9565	0.9156	0.9670	0.95958
0.98054	0.87226	0.8672	0.9356	0.9291	0.9735	0.94112
0.97405	0.88922	0.8752	0.9311	0.9096	0.9820	0.93413
0.97904	0.88772	0.8667	0.9341	0.9276	0.9730	0.94661
0.97255	0.88473	0.8687	0.9281	0.9191	0.9665	0.93413
0.97555	0.90519	0.8877	0.9466	0.9046	0.9810	0.93613
0.98004	0.91766	0.8852	0.9431	0.9496	0.9720	0.93114
0.97904	0.87375	0.8732	0.9316	0.9306	0.9540	0.91916
0.98204	0.88074	0.8742	0.9331	0.9111	0.9675	0.92914
0.97006	0.90419	0.8882	0.9535	0.9091	0.9725	0.94162
0.93762	0.90868	0.8827	0.9565	0.9156	0.9670	0.95958

In order to judge the performance of our fusion methods, Receiver Operating Character (ROC) curves were calculated for each of the network configurations for both the fused and non-fused datasets. Normally, ROC curves plotted are representative of the false positive rate against the true positive rate and as such they are often used for binary classifications performance measurements. The results of both ROC curves are shown in **Fig. 18 & 19**. For the non-fused dataset, every network configuration aside from VGG16 and 19 had roughly similar performances, and had AUC values above 0.9, however the fused dataset returned AUCs of 1.

Table 2. Fused Accuracy Results (DCNN)

AlexNet	VGG16	VGG19	Sum	Max	Min	Avg
1	0.99002	0.9650	0.999	0.9950	1	1
0.999	0.99202	0.9780	0.999	0.999	1	0.999
1	0.99102	0.9640	0.999	0.999	1	0.999
1	0.99102	0.9770	0.999	0.9950	0.9970	0.998
1	0.99002	0.9760	0.999	0.9940	1	0.999
1	0.99301	0.9790	1	0.999	1	0.998
0.998	0.99102	0.9660	0.9970	0.998	0.9960	0.999
0.999	0.99202	0.9590	0.998	0.9930	1	0.999
0.999	0.99401	0.9690	0.999	0.999	1	0.999
0.999	0.98703	0.9550	0.9970	0.9970	1	0.999

Table 3. Average Accuracy & Performance Times

	AlexNet	VGG16	VGG19	DCNN Fusion	CPU Time (Secs.)
Fused	0.87	0.96	0.95	0.952	79.4
Non-Fused	0.77	0.94	0.94	0.948	52.5

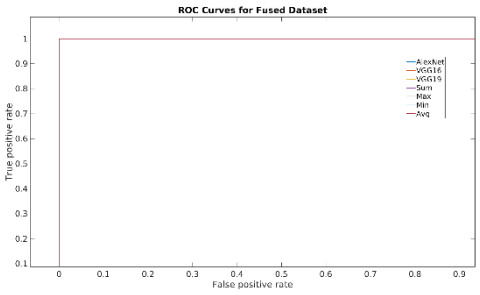


Fig. 18. ROC Curves for Fused Dataset

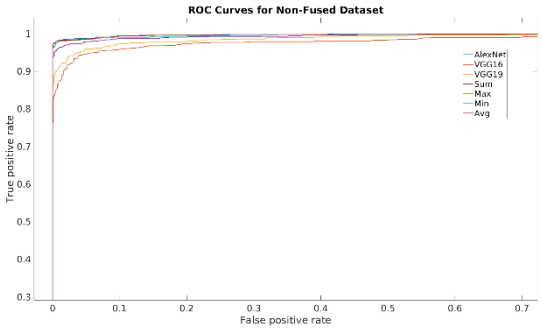


Fig. 19. ROC Curves for Non-Fused Dataset

Whenever a ROC curve results in an AUC of 1, immediate analysis must be done to determine the accuracy of the metric, as that would mean that the network classified every image correctly, which would be highly unlikely. Upon inspection of the classification results of the networks, it was determined that the ROC curve for the fused dataset was not accurate. However, the ROC and AUC for the non-fused dataset appeared to provide an accurate measure of the networks' performance.

5. Conclusion

The intent of this study was to investigate how image fusion can be used to improve the classification accuracy when using datasets that are comprised of images captured from different points of focus, modalities, or angles, in such a way that the dataset contains a substantial amount of overlapping or irrelevant information. The results of our study found that consolidating multiple corresponding images into one stream of data had the benefit of having equal if not better classification accuracy, while also requiring less time than the non-

fused dataset. Furthermore, utilizing our image fusion techniques in tandem with DCNN fusion along with an SVM classifier further increased the classification accuracy on average compared to the individual networks alone. In the future, we plan on expanding our image fusion technique to include a wider range of sensor types, modalities, etc. Additionally, future work involves further optimization of our techniques used to fuse images to include other multi-resolution transformations.

Acknowledgement

The intent of this study was to investigate how image fusion can be used to improve the classification accuracy when using datasets that are comprised of images captured from different points of focus, modalities, or angles, in such a way that the dataset contains a substantial amount of overlapping or irrelevant information. The results of our study found that consolidating multiple corresponding images into one stream of data had the benefit of having equal if not better classification accuracy, while also requiring less time than the non-fused dataset. Furthermore, utilizing our image fusion techniques in tandem with DCNN fusion along with an SVM classifier further increased the classification accuracy on average compared to the individual networks alone. In the future, we plan on expanding our image fusion technique to include a wider range of sensor types, modalities, etc. Additionally, future work involves further optimization of our techniques used to fuse images to include other multi-resolution transformations.

References

[1] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, “Context-based vision system for place and object recognition,” in *Proc. of Ninth IEEE International Conference on Computer Vision*, 2003. Article (CrossRef Link)

[2] A. Canziani, A. Paszke, and E. Culurciello, “An Analysis of Deep Neural Network Models for Practical Applications,” 2016. Article (CrossRef Link)

[3] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” *Neural Computation*, vol. 29, no. 9, pp. 2352-2449, 2017. Article (CrossRef Link)

[4] M. Giansiracusa, A. Lutz, S. Ezekiel, M. Alford, E. Blasch, A. Bubalo, and M. Thomas, "Multi-focus and multi-modal fusion: a study of multi-resolution transforms,” in *Proc. of SPIE 9841, Geospatial Informatics, Fusion, and Motion Video Analytics VI*, pp. 98410I, 2016. Article (CrossRef Link)

[5] M. Giansiracusa, A. Lutz, N. Messer, S. Ezekiel, E. Blasch, and M. Alford, “Bandelet- based image fusion: a comparative study for multi-focus images,” in *Proc. of Geospatial Informatics, Fusion, and Video Analytics VI, SPIE Defense + Security Conference, IEEE*, 2016. Article (CrossRef Link)

[6] A. Lutz, K. Grace, N. Messer, S. Ezekiel, E. Blasch, M. Alford, A. Bubalo, and M. Cornacchia. “Bandelet Transformation based Image Registration,” in *Proc. of Applied Image Pattern Recognition Workshop (AIPR), IEEE*, 2015. Article (CrossRef Link)

[7] A. Lutz, M. Giansiracusa, N. Messer, S. Ezekiel, E. Blasch, and M. Alford, “Optimal multi-focus contourlet-based image fusion algorithm selection,” in *Proc. of Geospatial Informatics, Fusion, and Video Analytics VI, SPIE Defense + Security Conference, IEEE*, 2016.

[8] Z. Omar and T. Stathaki, “Image Fusion: An Overview,” in *Proc. of 2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, pp. 306-310, 2014. Article (CrossRef Link)

[9] R. Gonzalez, R. Woods, *Digital Image Processing 3rd Edition*, Prentice Hall, 2008. https://www.pearson.com
[10] T. Stathaki, *Image Fusion: Algorithms and Applications*, Academic Press, 2008. https://www.researchgate.net/publication
[11] E. Blasch, E. Bossé, and D. Lambert, *High-Level Information Fusion Management and Systems Design*, Artech House, Norwood, MA, 2012. https://us.artechhouse.com
[12] E. Blasch, A. Steinberg, S. Das, J. Llinas, C. Chong, O. Kessler, E. Waltz, F. White, "Revisiting the JDL model for Information Exploitation," in *Proc. of Int'l Conf. on Info Fusion*, 2013. Article (CrossRef Link)
[13] K. Wang, H. Lin, P. Chan, S. Chang, "Application of Wavelet Decomposition and Gradient Variation in Texture Image Retrieval," in *Proc. of MUSP'08: Proceedings of the 8th WSEAS International Conference on Multimedia systems and signal processing*, pp. 299-304, 2008. Article (CrossRef Link)
[14] E. Candes, D. Donoho, "Ridgelets: A key to higher-dimensional intermittency?," *Philosophical Transactions of the Royal Society* A, vol. 357, no. 1760, pp. 2495- 2509, 1999. Article (CrossRef Link)
[15] E. Le Pennec, S. Mallat, "Bandelet image approximation and compression," *SIAM Journal of Multiscale Modeling and Simulation*, vol. 4, no. 3, pp. 992–1039, 2005. Article (CrossRef Link)
[16] S. Mallat, G. Peyre, "A review of Bandelet methods for geometrical image representation," *Numerical Algorithms*, vol. 44, pp. 205-234, 2007. Article (CrossRef Link)
[17] M. Do, M. Vetterli, "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation," *IEEE Transactions on Image Processing*, vol. 14 no. 12, pp. 2091- 2106, 2005. Article (CrossRef Link)
[18] J. Walker, *A Primer on Wavelets and their Scientific Applications 2nd Edition*, Chapman and HalUCRC, 2008. https://www.routledge.com
[19] J. Tarolli, "Multimodal Image Fusion with SIMS: Preprocessing with Image Registration," *Biointerphases*, vol. 11, no. 2, 14 January 2016. Article (CrossRef Link)
[20] N. Cvejic, D. Bull, and N. Canagarajah, "Region-Based Multimodal Image Fusion Using ICA Bases," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 743-751, 2007. Article (CrossRef Link)
[21] S. Li and B. Yang, "Hybrid Multiresolution Method for Multisensor Multimodal Image Fusion," *IEEE Sensors Journal*, vol. 10(9), pp. 1519-1526, 2010. Article (CrossRef Link)
[22] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective Assessment of Multi-Resolution Image Fusion Algorithms For Context Enhancement in Night Vision: A Comparative Study," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 34, No. 1, pp. 94-109, 2012. Article (CrossRef Link)
[23] G. Piella, "A General Framework for Multiresolution Image Fusion: From Pixels to Regions," *Information Fusion*, Vol. 4, Issue 4, pp. 259-280, December 2003. Article (CrossRef Link)
[24] F. Pedregosa, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825 –2830, 2011. Article (CrossRef Link)
[25] K. Xu, Z. Qin, G. Wang, H. Zhang, K. Huang and S. Ye, "Multi-focus Image Fusion using Fully Convolutional Two-stream Network for Visual Sensors," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 5, pp. 2253-2272, 2018. Article (CrossRef Link)
[26] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, J. Srivastava, "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection," in *Proc. of the 2003 SIAM International Conference on Data Mining*, pp. 25-36, 2003. Article (CrossRef Link)
[27] W. Huang, X. Wang, Y. Zhu and G. Zheng, "An improved kernel principal component analysis based on sparse representation for face recognition," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 6, pp. 2709-2729, 2016. Article (CrossRef Link)
[28] N. Kambhatla, T. Leen, "Dimension Reduction by Local Principal Component Analysis," *Neural computation*, vol. 9(7), pp. 1493-1516, 1997. Article (CrossRef Link)
[29] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern*, vol. 36, pp. 193–202, 1980.

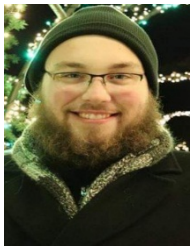
[30] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, L. Jackel, "Handwritten Digit Recognition with A Back-Propagation Network," *Advances in Neural Information Processing Systems 2*, pp. 396–404, 1990. Article (CrossRef Link)
[31] Y. LeCun, F. Huang, L. Bottou, "Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting," in *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004*, Vol. 2, pp. II-104, 2004. Article (CrossRef Link)
[32] H. Lee, R. Grosse, R. Ranganath, A. Ng, "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations," in *Proc. of the 26th Annual International Conference on Machine Learning. ACM*, pp. 609-616, 2009. Article (CrossRef Link)
[33] V. Vukotić, C. Raymond, and G. Gravier, "Multimodal and Crossmodal Representation Learning from Textual and Visual Features with Bidirectional Deep Neural Networks for Video Hyperlinking," in *Proc. of the 2016 ACM workshop on Vision and Language Integration Meets Multimedia Fusion*, pp. 37-44, 2016.
[34] D. Kornish, S. Ezekiel, and M. Cornacchia, "Fusion based Heterogeneous Convolutional Neural Networks Architecture," in *Proc. of 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2018. Article (CrossRef Link)
[35] N. Bodla, J. Zheng, H. Xu, J. Chen, C. Castillo, and R. Chellappa, "Deep Heterogeneous Feature Fusion for Template-Based Face Recognition," in *Proc. of 2017 IEEE Winter Conference on Applications of Computer IEEE*, 2017. Article (CrossRef Link)
[36] D. Kornish, S. Ezekiel, "DCNN Augmentation via Synthetic Data from Variational Autoencoders and Generative Adversarial Networks," in *Proc. of 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2018. Article (CrossRef Link)
[37] Petković, M., *Content-Based Video Retrieval*, Centre for Telematics and Information Technology, University of Twente, 2000. Article (CrossRef Link)
[38] McLaughlin, M. J., Lin, E-U., Ezekiel, S., et al., "Modified Deconvolution using Wavelet Image Fusion," in *Proc. of IEEE Applied Imagery Pattern Recognition Workshop*, 2014. Article (CrossRef Link)
[39] Blasch, E., Liu, Z., "LANDSAT Satellite Image Fusion Metric Assessment," in *Proc. of IEEE Nat. Aerospace Elec. Conf.*, 2011. Article (CrossRef Link)
[40] Anbumozhi, S., Manoharan. P.S., "Performance Analysis of High Efficient and Low Power Architecture for Fuzzy based image fusion," *American Journal of Applied Sciences*, Vol. 11, No. 5, pp. 769-781, 2014. Article (CrossRef Link)
[41] Deyun, C., Ming, G., Lei, S., Jianan, L., Yumei, Y., Li, W., "Image Fusion Method Based on Edge Feature Detection in Electrical Capacitance Tomography," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, pp. 267-274, 2014. Article (CrossRef Link)
[42] Ling, H., Wu, Y., et al., "Evaluation of Visual Tracking in Extremely Low Frame Rate Wide Area Motion Imagery," in *Proc. of Int'l. Conf. On Information Fusion*, 2011. Article (CrossRef Link)
[43] Blasch, E., Deignan, P. B. Jr, Dockstader, S. L., et al., "Contemporary Concerns in Geographical/Geospatial Information Systems (GIS) Processing," in *Proc. of IEEE Nat. Aerospace Electronics Conf (NAECON)*, 2011. Article (CrossRef Link)
[44] Blasch, E., Seetharaman, G., Palaniappan, K., Ling, H., Chen, G., "Wide-Area Motion Imagery (WAMI) Exploitation Tools for Enhanced Situation Awareness," in *Proc. of IEEE App. Imagery Pattern Rec. Workshop*, 2012. Article (CrossRef Link)
[45] Blasch, E., Liu, Z. Petkie, D., Ewing, R., et al., "Image Fusion of the Terahertz-Visual NAECON Grand Challenge Data," in *Proc. of IEEE National Aerospace and Electronics Conf. (NAECON)*, 2012. Article (CrossRef Link)
[46] Mei, X., Ling, H., Wu, Y., et al., "Efficient Minimum Error Bounded Particle Resampling L1 Tracker with Occlusion Detection," *IEEE Trans. on Image Processing (T-IP)*, Vol. 22, no. 7, pp. 2661 – 2675, 2013. Article (CrossRef Link)
[47] Blasch, E., Yang, C., et al., "Summary of Tracking and Identification Methods," *Proc. SPIE*, Vol. 9019, 2014. Article (CrossRef Link)

[48] Liang, P., et al., "Encoding Color Information for Visual Tracking: Algorithms and Benchmark," *IEEE Trans. on Image Processing*, vol. 24, no. 12, pp. 5630-5644, 2015. Article (CrossRef Link)
[49] Bin, W., Chao, W., "The Research of Remote Sensing Image Fusion Technology," *Applied Mechanics and Materials*, Vols. 513-517, pp. 3237-3240, 2014. Article (CrossRef Link)
[50] Piella, G., "A Region-Based Multiresolution Image Fusion Algorithm," in *Proc. of Int'l. Conf. On Information Fusion*, 2002. Article (CrossRef Link)

Abdullah A. Alshehri was born in 1964 in Saudi Arabia. He received his Bachelor of Science in Electrical Engineering at the University of Detroit. In 1999, he received his M.S. in Electrical Engineering followed by his Ph.D. in 2004 at the University of Pittsburgh. From 2005 to 2010, he was an assistant professor at the College of Telecom and Electronics and Jeddah College of Technology. Currently, he is an associate professor for the Electrical Engineering Department at King Abdulaziz University-Rabigh. His fields of interest consist of working with time-frequency, wavelet transform, neural networks, and statistical signal processing. Dr. Alshehri has been a member of IEEE since 1992, a member of the Saudi Engineers Council SEC since 2005 and has participated in multiple research projects at KAU.

Soundararajan Ezekiel received both his M.A. and Ph.D. from the University of Pittsburgh's department of mathematics. He also earned an M.S. degree at Loyola College in mathematics, post-graduate diploma in Operations Research at Anna University, and an MPhil from Madras Christian College in India. Currently, he is a computer science professor at Indiana University of Pennsylvania in the United States. Wavelet analysis, artificial intelligence, machine vision, deep learning, cyber security, and image and signal processing make up his areas of research. Dr. Ezekiel has received a three-time SFFP followship and seven-time VFRP fellowship.

Larry Pearlstein first earned a BSEE in 1982 from Drexel University, and, in 1987, a Ph.D. from Princeton University in electrical engineering. While serving as chairperson for the Video Specialists Group in the Advanced Television Systems Committee (ATSC), he led the effort to document the video compression portion of the ATSC Digital Television Standard. He also architected video subsystems for consumer electronics chips while working for ATI, AMD, and Broadcom. 71 US patents in digital television and consumer electronics belong to him. Currently, he works at The College of New Jersey as an associate professor of electrical and computer engineering and researches deep learning and connected devices.



John J. Conlen III is an undergraduate student at the Indiana University of Pennsylvania. He is majoring in Computer Science (Languages and Systems) and is taking a minor in Mathematics. During his time there he has participated in numerous research projects in fields such as Image Processing, machine learning, deep convolutional neural networks, recurrent neural networks, and blockchain.

Adam Lutz is a senior undergraduate researcher in Computer Science at Indiana University of Pennsylvania. Since 2016, he has collaborated with numerous organizations under his mentor including academic institutions and the AFRL to publish research in both domestic and international journals and conference proceedings. His published research fields include image & signal processing, wavelet analysis, deep learning, and IoT security.