PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY HIGHER EDUCATION AND SCIENTIFIC RESEARCH
KASDI MERBAH UNIVERSITY OUARGLA

Faculty of New Technologies of Information and Communication
Department of Electronics and Telecommunication

Thesis for Master Degree in
Automatic and Systems

Theme

Multi-focus image fusion with a deep convolutional neural network

Submitted by

Boudras Abdelwakil & Ammari Abderrahmane
Board of examiners:

Mr. Hamza azeddine MCB PresidentUKM Ouargla

Mme. louazene hassiba MCB Examinatrice UKM Ouargla

Mr. Benchabane abderrazak MCA Encadreur UKM Ouargla

Mr. Dida houdifa DoctorantCo-Encadreur UKM Ouargla

Academic year: 2020/ 2021

A b s t r a c t

As is well known, activity level measurement and fusion rule are two crucial factors in image fusion. For most existing fusion methods, either in spatial domain or in a transform domain like wavelet, the activity level measurement is essentially implemented by designing local filters to extract high-frequency details, or the calculated clarity information of different source images are then compared using some elaborately designed rules to obtain a clarity/focus map. Consequently, the focus map contains the integrated clarity information, which is of great significance to various image fusion issues, such as multi-focus image fusion, multi-modal image fusion, etc. However, in order to achieve a satisfactory fusion performance, these two tasks are usually difficult to finish. In this study, we address this problem with a deep learning approach, aiming to learn a direct mapping between source images and focus map. To this end, a deep convolutional neural network (CNN) trained by high-quality image patches and their blurred versions is adopted to encode the mapping. The main novelty of this idea is that the activity level measurement and fusion rule can be jointly generated through learning a CNN model, which overcomes the difficulty faced by the existing fusion methods. Based on the above idea, a new multi-focus image fusion method is primarily proposed in this paper. Experimental results demonstrate that the proposed method can obtain state-of-the-art fusion performance in terms of both visual quality and objective assessment. The computational speed of the proposed method using parallel computing is fast enough for practical usage. The potential of the learned CNN model for some other-type image fusion issues is also briefly exhibited in the experiments.

Keywords: multi-focus, image fusion, deep learning, convolutional neural network.

الـمـلـخص

كما هو معروف جيدًا ، يعد قياس مستوى النشاط وقاعدة الاندماج عاملين حاسمين في دمج الصورة. بالنسبة لمعظم طرق الاندماج الحالية ، سواء في المجال المكاني أو في مجال تحويل مثل المويجة ، يتم تنفيذ قياس مستوى النشاط بشكل أساسي من خلال تصميم المرشحات المحلية لاستخراج تفاصيل عالية التردد ، أو تتم مقارنة معلومات الوضوح المحسوبة لصور المصدر المختلفة باستخدام بعض قواعد مصممة بشكل متقن للحصول على خريطة الوضوح / التركيز. وبالتالي ، تحتوي خريطة التركيز البؤري على معلومات الوضوح المتكاملة ، والتي لها أهمية كبيرة في العديد من مشكلات دمج الصور ، مثل دمج الصور متعددة البؤرة ، ودمج الصور متعدد الوسائط ، وما إلى ذلك ، ومع ذلك ، من أجل تحقيق أداء اندماج مرضٍ ، فإن هذه عادة ما يكون من الصعب إنهاء مهمتين. في هذه الدراسة ، نعالج هذه المشكلة من خلال نهج التعلم العميق ، بهدف تعلم رسم الخرائط المباشر بين الصور المصدر وخريطة التركيز. تحقيقا لهذه الغاية ، تم اعتماد شبكة عصبية تلافيفية عميقة (CNN)

مدربة بواسطة تصحيحات صور عالية الجودة وإصداراتها غير الواضحة لتشفير التعيين. الحداثة الرئيسية لهذه الفكرة هي أنه يمكن إنشاء قياس مستوى النشاط وقاعدة الاندماج بشكل مشترك من خلال تعلم نموذج CNN ، والذي يتغلب على الصعوبة التي تواجهها طرق الاندماج الحالية. بناءً على الفكرة أعلاه ، تم اقتراح طريقة جديدة لدمج الصور متعددة التركيز بشكل أساسي في هذه الورقة. تظهر النتائج التجريبية أن الطريقة المقترحة يمكن أن تحصل على أحدث أداء اندماج من حيث الجودة البصرية والتقييم الموضوعي. السرعة الحسابية للطريقة المقترحة باستخدام الحوسبة المتوازية سريعة بما يكفي للاستخدام العملي. يتم أيضًا عرض إمكانات نموذج CNN الذي تم تعلمه لبعض مشكلات دمج الصور من النوع الآخر لفترة وجيزة في التجارب.
الكلمات المفتاحية: متعدد التركيز، صورة الانصهار، تعلم عميق، الشبكة العصبية التلافيفية.

Résumé

Comme on le sait, la mesure du niveau d'activité et la règle de fusion sont deux facteurs cruciaux dans la fusion d'images. Pour la plupart des méthodes de fusion existantes, que ce soit dans le domaine spatial ou dans un domaine de transformation comme l'ondelette, la mesure du niveau d'activité est essentiellement mise en œuvre en concevant des filtres locaux pour extraire des détails à haute fréquence, ou les informations de clarté calculées de différentes images sources sont ensuite comparées à l'aide de certains des règles élaborées pour obtenir une carte de clarté/focus. Par conséquent, la carte de mise au point contient les informations de clarté intégrées, ce qui est d'une grande importance pour divers problèmes de fusion d'images, tels que la fusion d'images multi-foyers, la fusion d'images multimodales, etc. Cependant, afin d'obtenir des performances de fusion satisfaisantes, ces deux tâches sont généralement difficiles à terminer. Dans cette étude, nous abordons ce problème avec une approche d'apprentissage en profondeur, visant à apprendre une correspondance directe entre les images sources et la carte de mise au point. À cette fin, un réseau de neurones à convolution profonde (CNN) formé par des patchs d'images de haute qualité et leurs versions floues est adopté pour coder la cartographie. La principale nouveauté de cette idée est que la mesure du niveau d'activité et la règle de fusion peuvent être générées conjointement par l'apprentissage d'un modèle CNN, ce qui surmonte la difficulté rencontrée par les méthodes de fusion existantes. Sur la base de l'idée ci-dessus, une nouvelle méthode de fusion d'images multi-foyers est principalement proposée dans cet article. Les résultats expérimentaux démontrent que la méthode proposée peut obtenir des performances de fusion de pointe en termes de qualité visuelle et d'évaluation objective. La vitesse de calcul de la méthode proposée utilisant le calcul parallèle est suffisamment rapide pour une utilisation pratique. Le potentiel du modèle CNN appris pour certains problèmes de fusion d'images d'un autre type est également brièvement exposé dans les expériences.
Mots clés : multi-focus, fusion d'images, deep learning, réseau de neurones convolutifs.

Acknowledgements

First of all, we thank ALLAH for giving us the strength to pass through the hard times that we have faced through our journey. We have experienced the guidance of Allah in each day of our lifetime and we will trust ALLAH as long as we are breathing. We would like to express our gratitude and appreciation to our advisor Mr.A.BENCHABANE and co-advisor Mr.H.DIDA for their academic guidance; they both were tremendous mentors to us. We would like to thank them for their encouragement and for encouraging us to seek for more knowledge because of their great devotion.

Also, we would like to thank our families' members especially our parents.

Our friends and colleagues for their support that helped us continue our study.

Finally, we are thankful to have the chance to study among a great family in Kasdi Merbah University of Ouargla, it was a pleasure.

Contents

List of Figures

List of Tables

List of abbreviations

CNN:convolutional neural network

IF:image fusion

VSN:visual sensor network

ANN:artificial neural network

FT: Fourier transform

IFT: Fourier transforms technique

DOF:depth-of-field

Introduction general

In the field of digital photography, it is often difficult for an imaging device like a digital single-lens reflex camera to take an image in which all the objects are captured in focus. Typically, un- der a certain focal setting of optical lens, only the objects within the depth-of-field (DOF) have sharp appearance in the photograph while other objects are likely to be blurred. A popular technique to obtain an all-in-focus image is fusing multiple images of the same scene taken with different focal settings, which is known as multi-focus image fusion. At the same time, multi-focus image fusion is also an important subfield of image fusion. With or without modification, many algorithms for merging multi-focus images can also be employed for other image fusion tasks such as visible-infrared image fusion and multi-modal medical image fusion (and vice versa). From this point of view, the meaning of studying multi-focus image fusion is twofold, which makes it an active topic in image processing community In recent years.In this paper, we address this problem with a deep learning approach, aiming to learn a direct mapping between source images and focus map. The focus map here indicates a pixel-level map which contains the clarity information after comparing the activity level measure of source images. To achieve this target, a deep convolutional neural network (CNN) [1] trained by high- quality image patches and their blurred versions is adapted to en- code the mapping. The main novelty of this idea is that the activity level measurement and fusion rule can be jointly generated through learning a CNN model, which overcomes the above difficulty faced by existing fusion methods. Based on this idea, we pro- pose a new multi-focus image fusion method in spatial domain. We demonstrate that the focus map obtained from the convolutional network is reliable that very simple consistency verification techniques can lead to high-quality fusion results. The computational speed of the proposed method using parallel computing is fast enough for practical usage. At last, we briefly exhibit the potential of the learned CNN model for some other-type image fusion issues, such as visible-infrared image fusion, medical image fusion and multi-exposure image fusion. To the best of our knowledge, this is the first time that the convolutional neural network is applied to an image fusion task. The most similar work was proposed by Li et al. [2]

Chapter 1: An Overview of Multi-Focus Image Fusion

I.1. Introduction

Currently there are many criteria that define the characteristics of a high-quality image such asSharp, correct color balance, correct exposure, focus, and lack thereof much noise has been suggested and discussed extensively in the image processing literature. Image quality has as much to do with user applications and a requirement as it does with perceived visual quality in general .

Imaging cameras, particularly those with long focal lengths, usually have a finite depth of field. In an image captured by those cameras, only those objects within the depth of field of the camera are focused, while other objects are blurred. Thus, the image that this camera gives is not clear to the human view in the part that is not in its depth of field. Image fusion is the process of combining useful and complementary information from multiple source images to create a single image .In this chapter we will present Image Fusion and Its different process and systems are used in different fusions, then we will discuss multi-focus image merging with its methods.

I. 2.Definition ofImage Fusion

The image fusion process is defined as gathering all the important information from multiple images and their inclusion into fewer images, usually a single one. This single image is more informative and accurate than any single source image, and it consists of all the necessary information. The purpose of image fusion is not only to reduce the amount of data but also to construct images that are more appropriate and understandable for the human and machine perception. [3]In computer vision,

multisensory image fusion is the process of combining relevant information from two or more images into a single image. [4] The resulting image will be more informative than any of the input images [5].

In remote sensing applications, the increasing availability of space borne sensors gives a motivation for different image fusion algorithms. Several situations in image processing require high spatial and high spectral resolution in a single image. Most of the available equipment is not capable of providing such data convincingly. Image fusion techniques allow the integration of different information sources. The fused image can have complementary spatial and spectral resolution characteristics. However, the standard image fusion techniques can distort the spectral information of the multispectral data while merging.

In satellite imaging, two types of images are available. The panchromatic image acquired by satellites is transmitted with the maximum resolution available and the multispectral data are transmitted with coarser resolution. This will usually be two or four times lower. At the receiver station, the panchromatic image is merged with the multispectral data to convey more information.Many methods exist to perform image fusion. The very basic one is the high pass filtering technique. Later techniques are based on Discrete Wavelet Transform, uniform rational filter bank, and Laplacian pyramid.

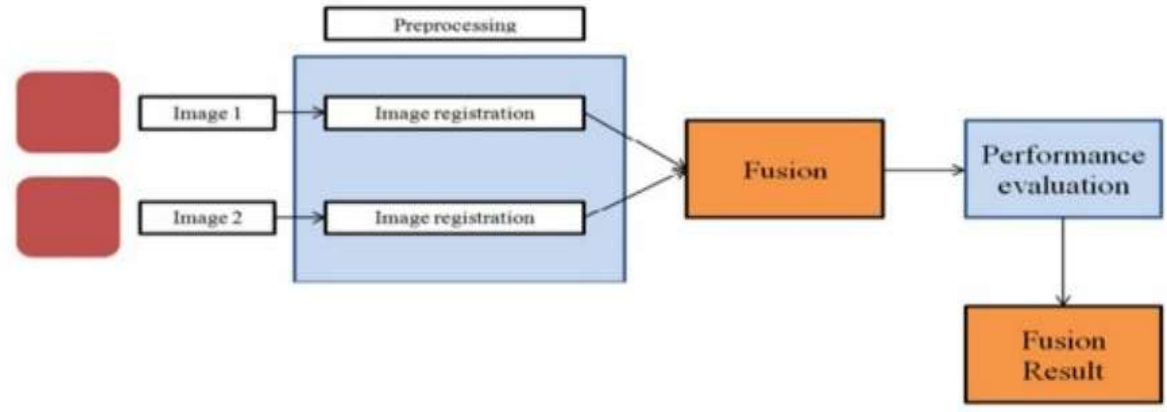


Fig (I.1): image fusion process.

I.3. Image Fusion Process

As discussed earlier, the goal of IF is to produce a merged image with the integration of Information from more than one image. **Fig (I.2)** demonstrates the major steps involved in IF process. In wide-ranging, the registration is measured as an optimization issue which is used to exploit the similarity as well as to reduce the cost. The Image registration procedure is used to align the subsequent features of

various images with respect to a reference image. In this procedure, multiple source images are used for registration in which the original image is recognized as a reference image and the original images are aligned through reference image. In feature extraction, the significant features of registered images are extracted to produce several feature maps.

By employing a decision operator whose main objective is to label the registered images with respect to pixel or feature maps, a set of decision maps are produced. Semantic equivalence obtained the decision or feature maps that might not pass on to a similar object. It is employed to connect these maps to a common object to perform fusion.

This process is redundant for the source obtained from a similar kind of sensors. Then, radiometric calibration is employed on spatially aligned images. Afterward, the transformation of feature maps is performed on an ordinary scale to get end result in a similar representation format.

Finally, IF merge the consequential images into a one resultant image containing an enhanced explanation of the image. The main goal of fusion is getting more informative fused image [6].

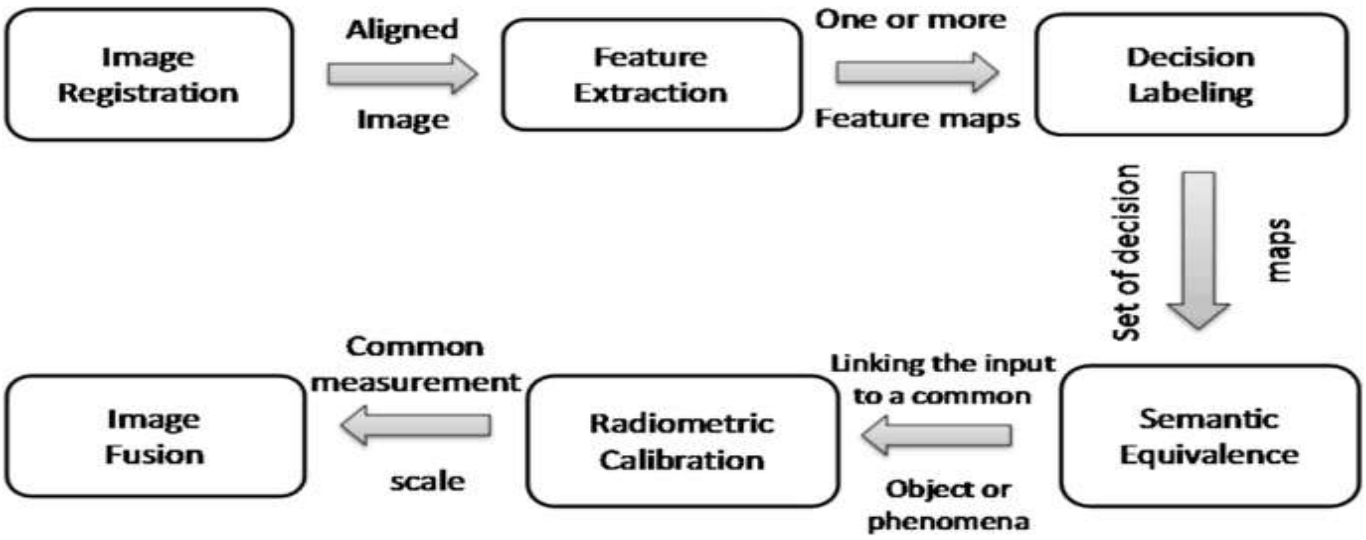


Fig (I.2): The main steps of Image Fusion procedure

I. 3.1.Steps process in Image Fusion

Fundamental steps involved in image fusion process are shown in **Fig (I.3)**. It consists of 5 major steps:

1. Pre-processing,
2. Image registration,
3. Image fusion,
4. Post-processing and

5. Fusion performance evaluation.

- **Step1**

In pre-processing stage, noise or artefacts introduced in the source images during image acquisition process are removed or reduced.

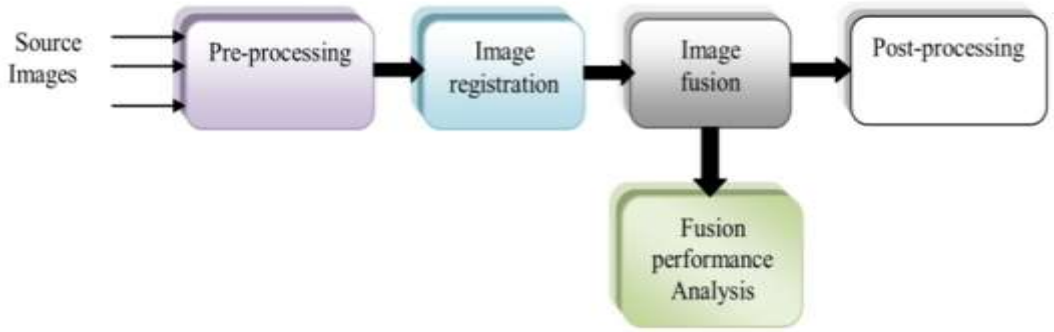


Fig (I.3): Steps process in Image Fusion

- **Step2**

Image registration is the process of aligning or arranging more than one images of a same scene according to a co-ordinate system. In this process, one of the source images will be taken as a reference image. It is also termed as the fixed image. Then geometric transformation will be applied on remaining source images to align them with the reference image.

- **Step3**

Fusion process can be performed at three levels (Ardeshir and Nikolov, 2007): pixel, feature and decision. Pixel level fusion is done on each input image pixel by pixel. However at feature level, fusion is executed on the extracted features of source images. At decision level, fusion is performed on probabilistic decision information of local decision makers. These decision makers are in turn derived from the extracted features. Pixel level fusion schemes are preferable for fusion compared to other level approaches because of their effectiveness and ease of implementation. In this thesis, our interest is only on pixel level fusion schemes.

- **Step4**

During the fusion process, some required information of source images may be lost and visually unnecessary information or artifacts may be introduced into the fused image. Hence, fusion algorithms need to be assessed and evaluated for better performance. This performance analysis can be carried out by evaluating them qualitatively by visual inspection and quantitatively using fusion metrics.

- **Step5**

In post-processing, fused images are further processed depending on the application. This processing may involve segmentation, classification and feature extraction. In this thesis, we developed new pixel level image fusion algorithms for both multi-focus and multi-modal images.

I. 4.Image fusion systems

- **Single Sensor**

Single sensor captures the real world as a sequence of images. The set of images are fused together to generate a new image with optimum information content. For example in illumination variant and noise full environment, a human operators like detector operator may not be able to detect objects of his interest which can be highlighted in the resultant fused image.

The shortcoming of this type of systems lies behind the limitations of the imaging sensor that are being used in other sensing area. Under the conditions in which the system can operate, its dynamic range, resolution, etc. are all restricted by the competency of the sensor. For example, a visible-band sensor such as the digital camera is appropriate for a brightly illuminated environment such as daylight scenes but is not suitable for poorly illuminated situations found during night time, or under not good conditions such as in fog or rain.

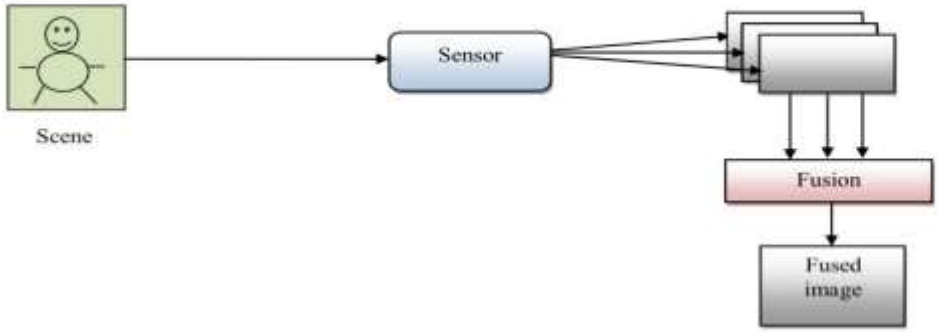


Fig (I.4): Single Sensor system

- **Multi Sensor**

A multi-sensor image fusion scheme overcomes the limitations of a single sensor image fusion by merging the images from several sensors to form a composite image an infrared camera is accompanying the digital camera and their individual images are merged to obtain a fused image. This approach

overcomes the issues referred to before. The digital camera is suitable for daylight scenes; the infrared camera is appropriate in poorly illuminated environments. It is used in military area, machine vision like in object detection, robotics, medical imaging. It is used to solve the merge information of the several images.

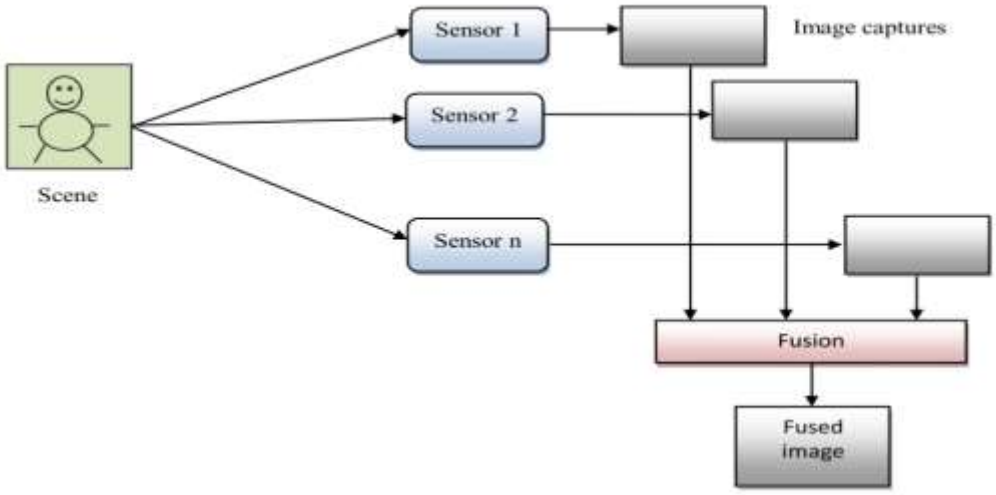


Fig (I.5): Multi-sensor system.

- **Multiview Fusion**

In this images have multiple or different views at the same time. Multimodal Fusion: Images from different models like panchromatic, multispectral, visible, infrared, remote sensing. Common methods of image fusion

• Weighted averaging pixel wise

• Fusion in transform domain

• Object level fusion

- **Multifocus Fusion**

Images from 3d views with its focal length The original image can be divided into regions such that every region is in focus in at least one channel of the image.

I. 5. Image Fusion Techniques

IF techniques can be classify as spatial and frequency domainThe spatial technique deals with pixel values of the input images in which the pixels values are manipulated to attain a suitable outcome. The entire synthesis operations are evaluated using Fourier Transform (FT) of the image and then IFT is evaluated to obtain a resulting image. Other IF techniques are PCA, IHS and high pass filtering and brovey method [7]. Discrete transform fusion techniques are extensively used in image fusion as compared to pyramid based fusion technique. Different types of IF techniques are shown in **Fig (I.6)** [8].

I.5.1. Spatial Based Techniques

The Spatial based technique is a simple image fusion method consist of Max–Min, Minimum, Maximum, Simple Average and Simple Block Replace techniques [9]. Table 1shows the diverse spatial domain based methods with their pros and cons.

- **Simple Average**

It is a fusion technique used to combined images by averaging the pixels. This technique focused on all regions of the image and if the images are taken from the same type of sensor then it works well [10]. If the images have high brightness and high contrast then it will produce good results.

- **Minimum Technique**

It selects the lowest intensity value of the pixels from images and produced fused image [9]. It is used for darker images.

- **Maximum Technique**

It selects the pixels values of high intensity from images to yield fused image [7].

- **Max–Min Technique**

It selects the averaging values of the pixels smallest and largest from the entire source images and produced the resultant merged image.

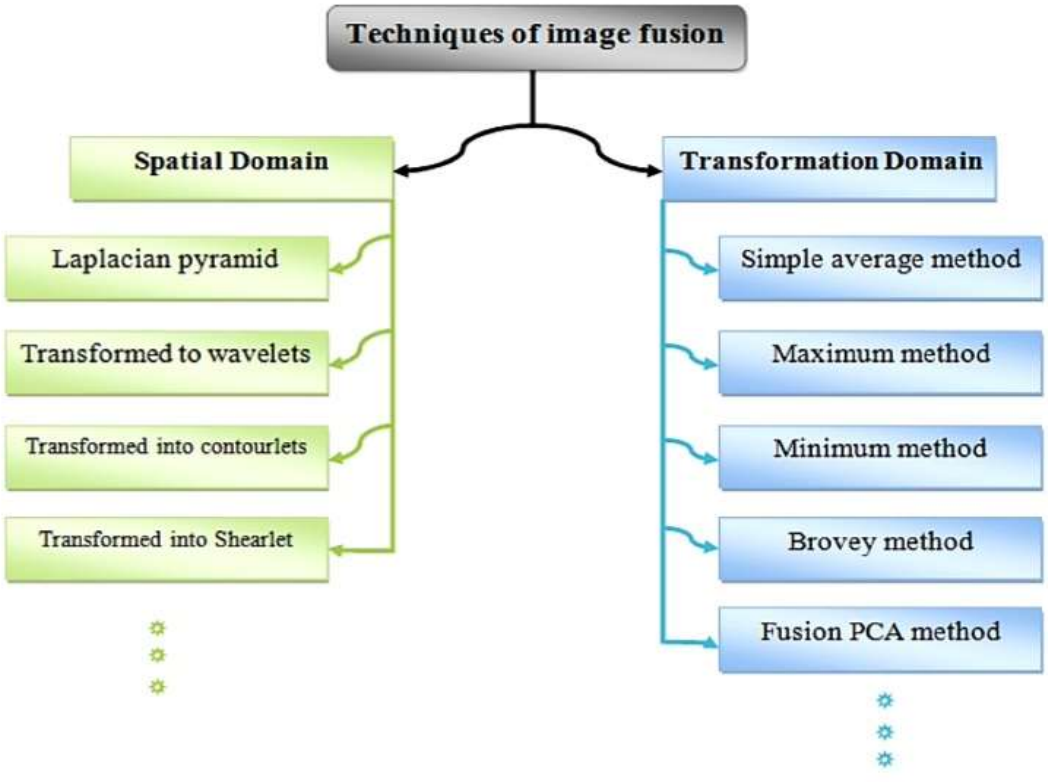


Fig (I.6): Image Fusion Techniques.

I.6.Classification of Image Fusion Methods

Image fusion process can be divided into three categories:

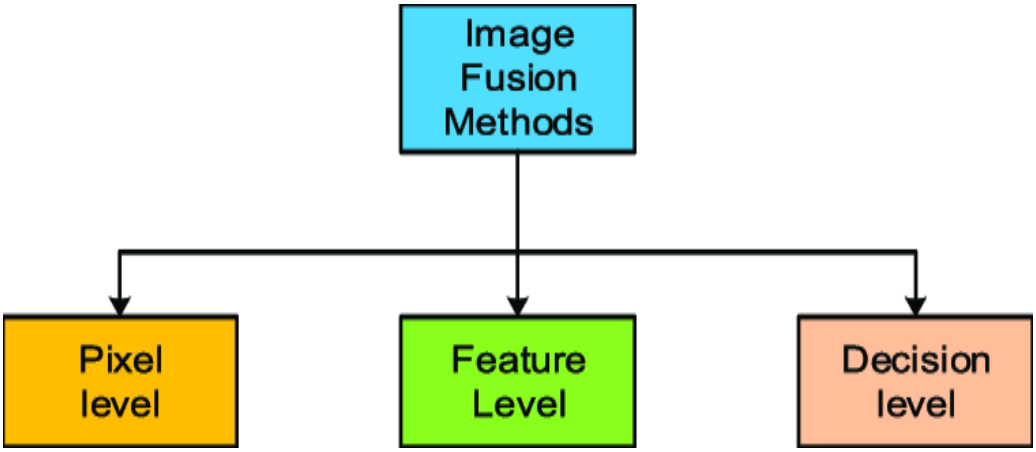


Fig (I.7): Classification of Image Fusion Methods

- **Pixel-level**

Pixel-level is the lowest level of the image fusion process. It deals directly with the pixels. The advantages of this level are to detect unwanted noise, to provide detail information, less complexity, and ease of implementation. However, these methods do not handle mis-registration and can cause blocking artefact.

- **Feature-level**

In feature level process, features are extracted from input images. Image is segmented in continuous regions and fuse them using fusion rule. Features of images are combined such as size, shape, contrast, pixel intensities, edge and texture.

- **Decision-level**

Decision level consists of merging information at a higher level of abstraction, combines the results from multiple algorithms to yield a final fused decision.

I.7.Application and Use of Image Fusion

1. Fusion is basically used remote or satellite area for the proper view of satellite vision
2. It must use in medical imaging where disease should analyses through imaging vision through spatial resolution and frequency perspectives.

3. Image fusion used in military areas where all the perspectives used to detect the threats and other resolution work based performance.
4. For machine vision it is effectively used to visualize the two states after the image conclude its perfect for the human vision.
5. in robotics field fused images mostly used to analyse the frequency variations in the view of images.
6. Image fusion is used in artificial neural networks in 3d where focal length varies according to wavelength transformation.

I.8.Adventages and Disadvantages Image Fusion:

- **Advantages**

1. It is easiest to interpret.
2. Fused image is true in color.
3. It is best for identification and recognition
4. It is low in cost
5. It has a high resolution used at multiscale images.
6. Through image fusion there is improved fused images in fog
7. Image fusion maintains ability to read out signs in all fields.
8. Image fusion has so many contrast advantages basically it should enhance the image with all the perspectives of image.
9. It increases the situational or conditional awareness.
10. Image fusion reduced the data storage and data transmission.

- **Disadvantages**

1. Images have less capability in adverse weather conditions it is commonly occurred when image fusion is done by single sensor fusion technique.
2. Not easily visible at night it is mainly due to camera aspects whether it is in day or night.
3. More source energy is necessary for the good

I.9. Multi Focus Image Fusion

Multi-focus image fusion is used to collect useful and necessary information from input images with different focus depths in order to create an output image that ideally has all information from input images In visual sensor network (VSN), sensors are cameras which record images and video sequences. In many applications of VSN, a camera can't give a perfect illustration including all details of the scene. This is because of the limited depth of focus exists in the optical lens of cameras. Therefore, just the

object located in the focal length of camera is focused and cleared and the other parts of image are blurred. VSN has an ability to capture images with different depth of focuses in the scene using several

cameras. Due to the large amount of data generated by camera compared to other sensors such as pressure and temperature sensors and some limitation such as limited band width, energy consumption and processing time, it is essential to process the local input images to decrease the amount of transmission data. The aforementioned reasons emphasize the necessary of multi-focus images fusion. Multi-focus image fusion is a process which combines the input multi-focus images into a single image including all important information of the input images and it's more accurate explanation of the scene than every single input image [6].

Fig (I.8): Multi Focus Image Fusion

I.10.Objectives evaluation metrics

In this work we will use the statistics of Petrovic to provide informationmuch more detailed on the advantages and disadvantages of the fusion method inestimating the information contribution of each image source such as [11]:

Total fusion performance**QXY / F**, fusion loss**LXY / F**, andArtefacts de lafusion**NXY / F**.

- **Total performance of the fusion QXY / F**

Consider two source images X, Y and a merged image$F$. The total performance of the fusion QXY / F of the source and merged images of size $M \times N$ is calculated by [9]:

$$Q^{XF/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}(Q^{XF}(n,m)W^{X}(n,m)+Q^{YF}(n,m)W^{Y}(n,m)}{\sum_{n=1}^{N}\sum_{m=1}^{M}(W^{X}(n,m)+W^{Y}(n,m))} \quad \textbf{(I.1)}$$

Où

WX et WY sont des pondérations attribuées à QXF et QYF respectivement.$QXFetQYF$Sont les préservations des informations de bord des images sources$XetY$ respectivement et qui sont calculées comme suit:

$$Q^{XF}(n,m) = Q_g^{XF}(n,m)Q_\alpha^{XF}(n,m)$$
$$Q^{YF}(n,m) = Q_g^{YF}(n,m)Q_\alpha^{YF}(n,m) \quad \textbf{(I.2)}$$

Avec

$$Q_g^{XF}(n,m) = \frac{\gamma_g}{1 + e^{K_g}(G^{XF}(n,m) - \sigma_g}$$

$$Q_\alpha^{XF}(n,m) = \frac{\gamma_g}{1+e^{K_\alpha}(X^{XF}(n,m)-\sigma_\alpha}} \quad \textbf{(I.3)}$$

avec

$$G^{XF}(n,m) = \begin{cases} \frac{g_F(n,m)}{g_X(n,m)} \; si \; g_x(n,m) > g_F(n,m) \\ \frac{g_X(n,m)}{g_F(n,m)} \; ailleur \end{cases}$$

$$g_{x=}\sqrt{s_X^x(n,m)^2 + s_X^y(n,m)^2} \quad \textbf{(I.4)}$$

$$X^{XF}(n,m) = 1 - \frac{|a_X(n,m) - a_F(n,m)|}{\frac{\pi}{2}} \quad \textbf{(I.5)}$$

$$a_{x=}arctang(\frac{S_X^y(n,m)}{S_X^x(n,m)})$$

The constants $\gamma g, Kg, K\alpha, \sigma\alpha$ andσg are used to estimate the shape of the sigmoids usedto determine the edge and orientation.

The total performance of the fusion satisfies [11]:$\mathbf{0 \leq QXY / F \leq 1}$

- If QXY / F = 0 then this implies a complete loss of the source information.

- If $\boldsymbol{QXY / F} = 1$ then indicates the ideal fusion 'without loss of source information.

- **Loss of fusion$\boldsymbol{LXY / F}$:**

It measures edge information lost in the merge job. This information is not shown in the merged image but in the source images. The mathematical expressionis given by [9]:

$$L^{XF/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} r(n,m)\left(\left(1 - Q^{XF}(n,m)\right)W^{X}(n,m) + (1 - Q^{YF}(n,m))W^{Y}(n,m)\right)}{\sum_{n=1}^{N}\sum_{m=1}^{M}(W^{X}(n,m) + W^{Y}(n,m))}$$

$$\text{Ou } r(n,m) = \begin{cases} 1 \; si \; g_F(n,m) < g_x(n,m) \;\; ou \; g_F(n,m) < g_y(n,m) \\ 0 \qquad\qquad ailleur \end{cases} \qquad \textbf{(I.6)}$$

The melting loss range is **0 ≤ LXY / F ≤ 1**:

- If $\boldsymbol{LXY / F} = 0$ the fusion loss is low.
- If $\boldsymbol{LXY / F} = 1$ the fusion loss is high.

It is necessary that the value of **LXY / F** lower to obtain a better performance of anyfusion algorithm.

- **Artefacts de la fusion $\boldsymbol{NXY / F}$**

Unnecessary visual information can be introduced into the combined image which does notany relevance to the source images. These artifacts should be avoided. The expressionmathematics of **NXY / F** is given by [11]:

$$N_K^{XY/F} = \frac{\sum_n \sum_m AM_{n,m}\left((1 - Q^{XF}(n,m))W^{X}(n,m) + (1 - Q^{YF}(n,m))W^{Y}(n,m)\right)}{\sum_n \sum_m \left(W^{X}(n,m) + W^{Y}(n,m)\right)}$$

$$\text{Où} \quad AM_{n,m} = \begin{cases} 1 \; si \; g_F(n,m) > g_x(n,m) \; et \; g_F(n,m) > g_y(n,m) \\ 0 \qquad ailleur \end{cases} \qquad \textbf{(I.7)}$$

This fusion metric satisfies **0 ≤ NXY / F ≤ 1**.

- If **NXY / F** = 0 then there is no fusion artefacts.
- If **NXY / F** = 1 then there is a serious degradation of the image quality because of the noise or artifacts.

We see that the information score on fusion, loss and fusion artefacts arecomplementary to each other. The sum of these measures must give the unit:

$$Q^{XY/F} + L^{XY/F} + N^{XY/F} = 1 \quad \textbf{(I.8)}$$

I.11.Conclusion:

Image fusion is a complex process because it is made up of many interdependent phases, which makes it interdependent phases, which makes it a technique that is all the more difficult to grasp as there is no universal fusion procedure. The quality of the fusion images presented to the clinician will depend on the choice of techniques associated with each step of the fusion; these associations are therefore intended to give informative images for the diagnosis. In this chapter, we presented multi-focus image fusion with its different methods, and then we discussed image fusion and its different approaches and Methods used for fusion.

Chapter 2: Deep learning Convolutional Neural Network for Image Fusion

II.1. Introduction

Convolutional Neural Network has had ground breaking results over the past decade in a variety of fields related to pattern recognition; from image processing to voice recognition. The most beneficial aspect of CNNs is reducing the number of parameters in ANN. This achievement has prompted both researchers and developers to approach larger models in order to solve complex tasks, which was not possible with classic ANNs; The most important assumption about problems that are solved by CNN should not have features which are spatially dependent. In other words, for example, in a face detection application, we do not need to pay attention to where the faces are located in the images. The only concern is to detect them regardless of their position in the given images. Another important aspect of CNN is to obtain abstract features when input propagates toward the deeper layers. For example, in image classification, the edge might be detected in the first layers, and then the simpler shapes in the second layers, and then the higher level features such as faces in the next layers.In this chapter we will present Concepts of Convolutional Neural Network and CNN model for image fusion.

II.2. Concepts of Convolutional Neural Network

Convolutional Neural Network (CNN), also called ConvNet, is a type of Artificial Neural Network(ANN), which has deep feed-forward architecture and has amazing generalizing ability as compared to other networks with FC layers, it can learn highly abstracted features of objects especially spatial data and can identify them more Efficiently.

A deep CNN model consists of a Finite set of processing layers that can learn various features of input data (e.g. image) with multiple level of abstraction. The initiatory layers learn and extract the high

level features (with lower abstraction), and the deeper layers learns and extracts the low level features (with higher abstraction). The basic conceptual model of CNN was shown in **Fig (II.1)**[12], different types of layers described in subsequent sections.

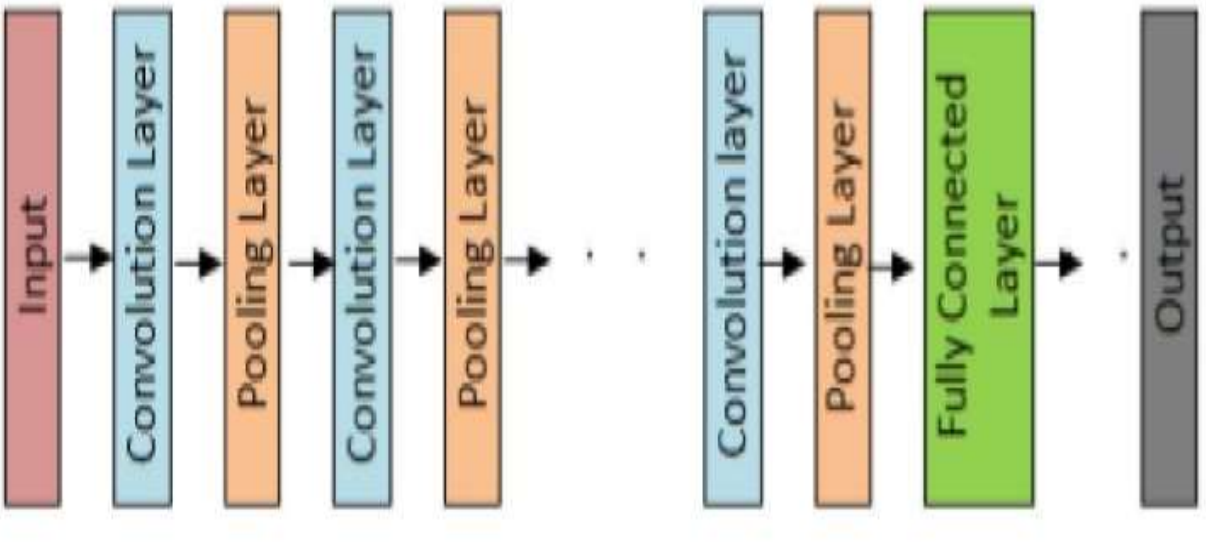


Fig (II.1): Conceptual model of CNN.

Why Convolutional Neural Networks is more considerable over other classical neural networks in the context of computer vision?

- One of the main reason for considering CNN in such case is the weight sharing feature of CNN, that reduce the number of trainable parameters in the network, which helped the model to avoid over Fitting and as well as to improved generalization.
- In CNN, the classification layer and the feature extraction layers learn together, that makes the output of the model more organized and makes the output more dependent to the extracted features.
- The implementation of a large network is more difficult by using other types of neural networks rather than using Convolutional Neural Networks.

Nowadays CNN has been emerged as a mechanism for achieving promising result in various computer vision based applications like image classification, object detection, face detection, speech recognition, vehicle recognition, facial expression recognition, text recognition and many more.

Now description of different components or basic building blocks of CNN briefly as follows.

II.3Network Layers

As we mentioned earlier, that a CNN is composed of multiple building blocks (known as layers of the architecture), in this subsection, we described some of these building blocks in detail with their role in the CNN architecture.

II .3.1 Convolutional Layer

Convolutional layer1 is the most important component of any CNN architecture. It contains a set of convolutional kernels (also called Filters), which gets convolved with the input image (N-dimensional metrics) to generate an output feature map.

- **What is a kernel?**

A kernel can be described as a grid of discrete values or numbers, where each value is known as the weight of this kernel. During the starting of training process of an CNN model, all the weights of a kernel are assigned with random numbers (different approaches are also available there for initializing the weights). Then, with each training epoch, the weights are tuned and the kernel learned to extract meaningful features. In **Fig (II.2)**, we have shown 2D Filter.

0	1
-1	2

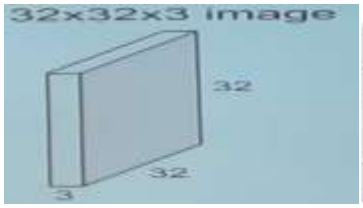
Fig (II.2): Example of a 2 x 2 kernel.

- **What is Convolution Operation?**

Before we go any deeper, let us first understand the input format to CNN.

Unlike other classical neural networks (where the input is in a vector format), in CNN the input is a multi-channelled image (e.g. for RGB image as in **Fig (II.3)**, it is 3 channelled and for Gray-Scale image, it is single channelled)[12].

Now, to understand the convolution operation, if we take a gray-scale image of 4 x 4 dimensions, shown in **Fig (II.4)**and a 2 x 2 kernels with randomly initialized weights as shown in **Fig (II.5)**.



1	0	-2	1
-1	0	1	2
0	2	1	0
1	0	0	1

Fig (II.3): Example of a RGB image**Fig (II.4):**4x4 Gray-Scale image

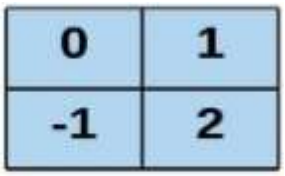


Fig (II.5): kernel of size 2x2.

Now, in convolution operation, we take the 2 x 2 kernel and slide it over all the complete 4 x 4 image horizontally as well as vertically and along the way we take the dot product between kernel and input image by multiplying the corresponding values of them and sum up all values to generate one scaler value in the output feature map. This process continues until the kernel can no longer slide further.

To understand the thing more clearly, let's do some initial computations performed at each step graphically as shown in **Fig (II.6)**, where the 2 x 2 kernel (shown in light blue color) is multiplied with the same sized region (shown in yellow color) within the 4 x 4 input image and the resulting values are summed up to obtain a corresponding entry (shown in deep blue) in the output feature map at each convolution step[16].

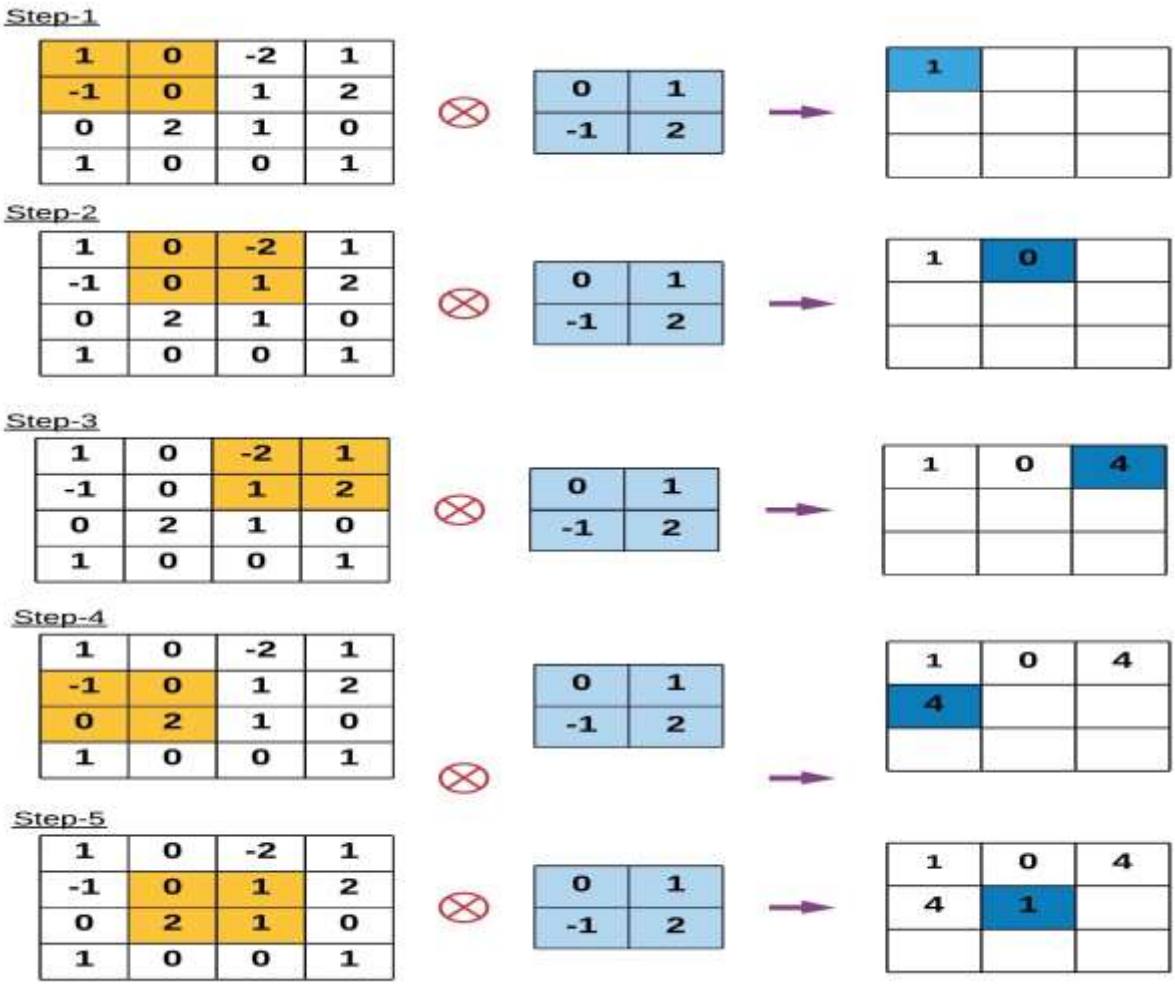


Fig (II.6): Illustrating the first 5 steps of convolution operation

After performing the complete convolution operation, the final output feature map is shown in **Fig (II.7)**as follows.

1	0	4
4	1	1
1	1	2

Fig (II.7): The final feature map after the complete convolution operation

In the above example, we apply the convolution operation with no padding to the input image and with stride (i.e. the taken step size along the horizontal or vertical position) of 1 to the kernel. But we can use other stride value (rather than 1) in convolution operation.

The noticeable thing is if we increase the stride of the convolution operation, it resulted in lower-dimensional feature map.

The padding is important to give border size information of the input image more importance, otherwise without using any padding the border side features are gets washed away too quickly. The padding is also used to increase the input image size, as a result the output feature map size also get increased. The **Fig (II.8)** gives an example by showing the convolution operation with Zero-padding and 3 stride value.

The formula to find the output feature map size after convolution operation as below:

$$h' = \left\lfloor \frac{h - f}{s} \right\rfloor$$
$$w' = \left\lfloor \frac{w - f}{s} \right\rfloor$$

Where **h'** denotes the height of the output feature map, **w'** denotes the width of the output feature map, **h** denotes the height of the input image, **w** denotes the width of the input image, **f** is the filter size, **p** denotes the padding of convolution operation and **s** denotes the stride of convolution operation.

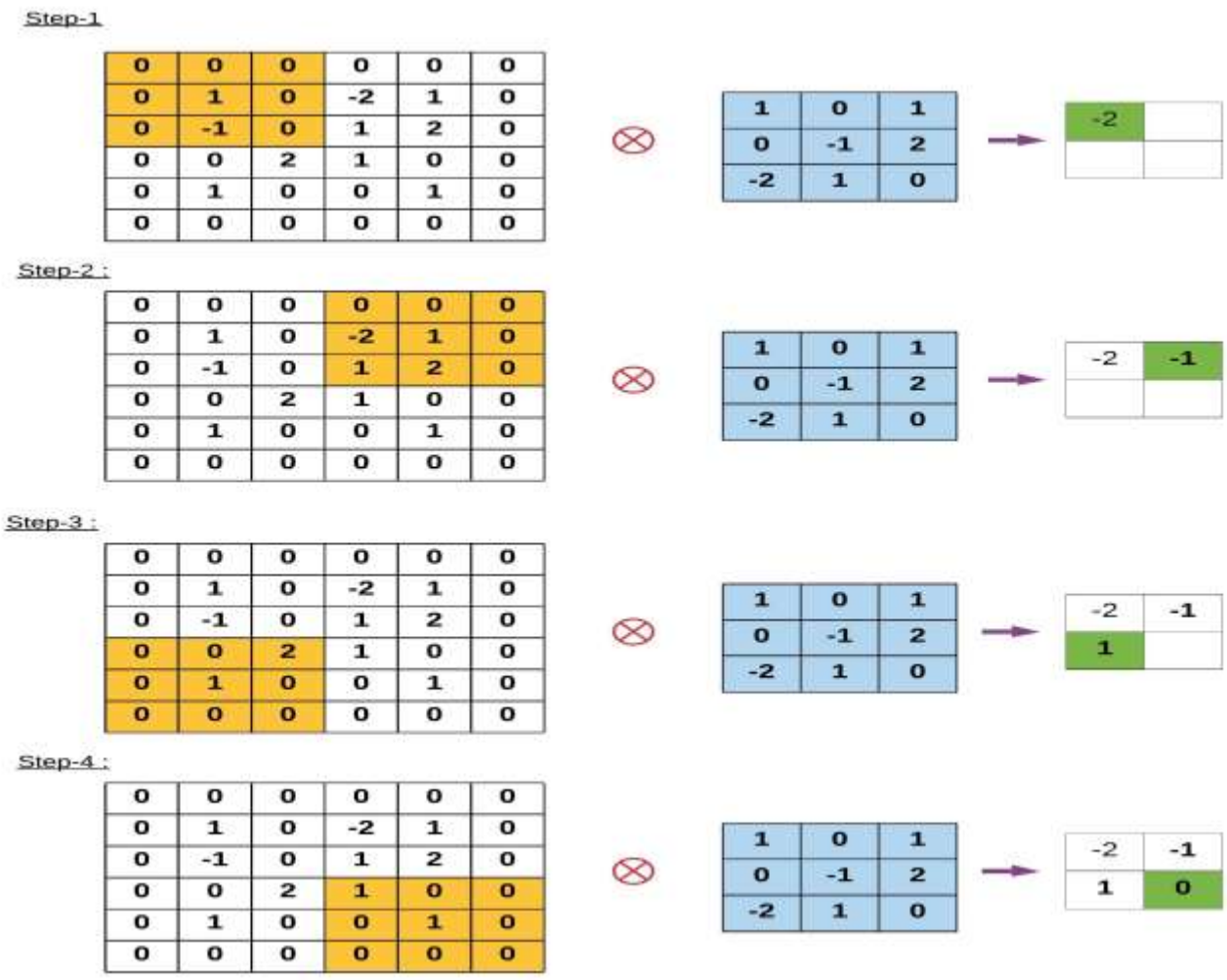


Fig (II.8): The computations performed at each step, where the 3 x 3 kernel (shown in light blue color) is multiplied with the same sized region (shown in yellow color) within the 6 x 6 input image (where we applied zero-padding to the original input image of 4 x 4 dimension and it becomes of 6 x 6 dimensional) and values are summed up to obtain a corresponding entry (shownin deep green) in the output feature map at each convolution step.

II.4 Pooling Layer

The pooling layers are used to sub-sample the feature maps (produced after convolution operations), i.e. it takes the larger size feature maps and shrinks them to lower sized feature maps.

While shrinking the feature maps it always preserve the most dominant features (or information) in each pool steps. The pooling operation is performed by specifying the pooled region size and the stride of the operation, similar to convolution operation. There are different types of pooling techniques are used in deferent pooling layers such as max pooling, min pooling, average pooling, gated pooling, tree pooling, etc. Max Pooling is the most popular and mostly used pooling technique[12].

The main drawback of pooling layer is that it sometimes decreases the overall performance ofCNN. The reason behind this is that pooling layer helps CNN to find whether a specific featureis present in the given input image or not without caring about the correct position of thatfeature.

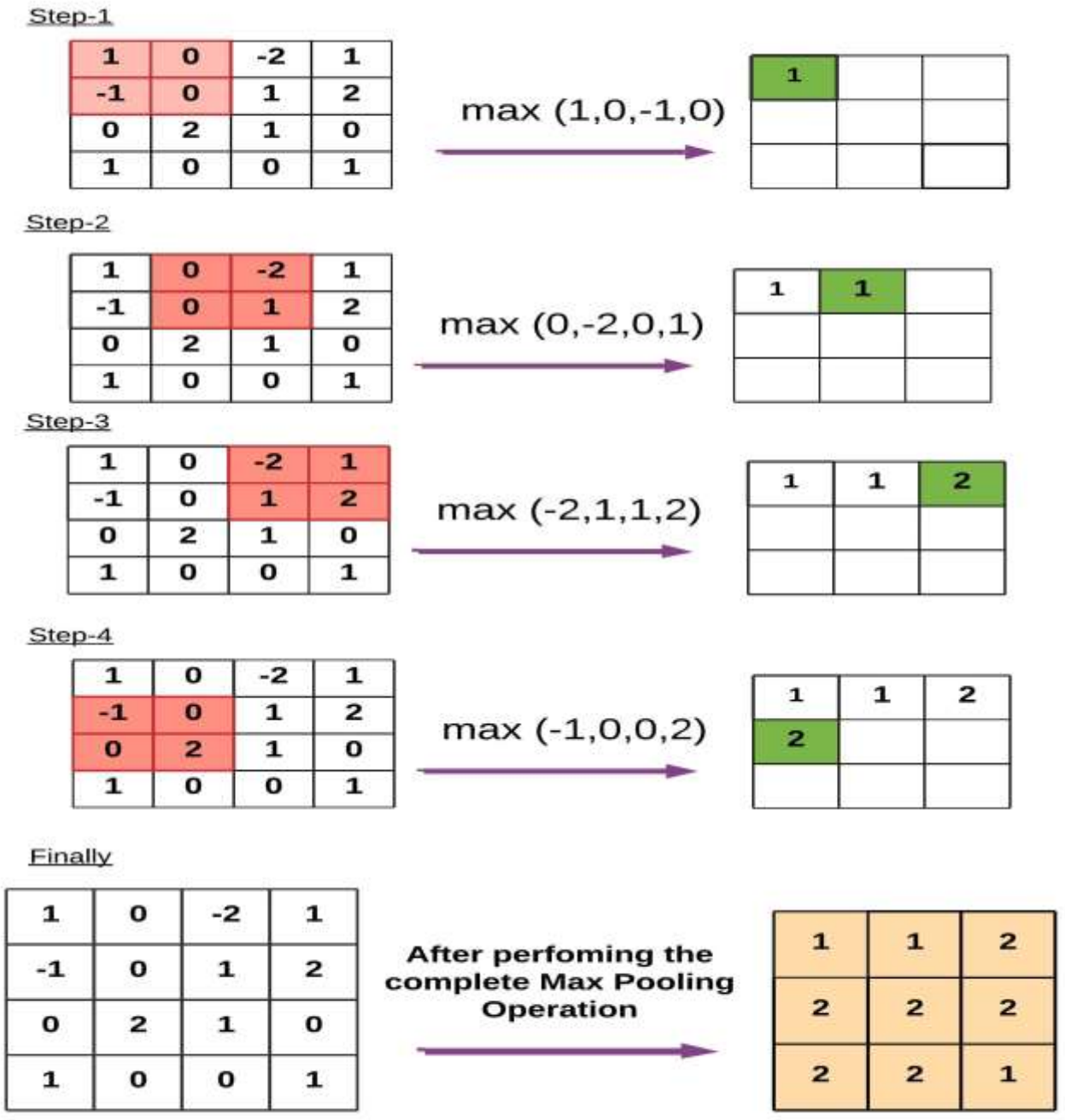


Fig (II.9): Illustrating an example that shows some initial steps as well as the final output ofmax-pooling operation, where the size of the pooling region is 2 x 2 (shown in orange color, in the input feature map) and the stride is 1 and the corresponding computed value in the output feature map (shown in green)

II.5. CNN model for image fusion

- **CNN model**

CNN is a typical deep learning model, which attempts to learn a hierarchical feature representation mechanism for signal/image data with different levels of abstraction. More concretely, CNN is a trainable multi-stage feed-forward artificial neural network and each stage contains a certain number of feature maps corresponding to a level of abstraction for features. Each unit or coefficient in a feature map is called a neuron. The operations such as linear convolution, non-linear activation and spatial pooling applied to neurons are used to connect the feature maps at different stages.

Local receptive fields, shared weights and sub-sampling are three basic architectural ideas of CNNs [1]. The first one indicates a neuron at a certain stage is only connected with a few spatially neigh boring neurons at its previous stage, which is in accord with the mechanism of mammal visual cortex. As a result, local convolutional operation is performed on the input neurons in CNNs, unlike the fully-connected mechanism used in conventional multilayer perception. The second idea means the weights of a convolutional kernel is spatially invariant in feature maps at a certain stage. By combining these two ideas, the number of weights to be trained is greatly reduced. Mathematically, let x i and y j denote the i -th input feature map and j -th output feature map of a convolutional layer, respectively. The 3D convolution and non-linear ReLU activation [13] applied in CNNs are jointly expressed as:

$$y^j = \max(0, b^j + \sum_i k^{ij} * x^i), \tag{II.1}$$

Where k ij is the convolutional kernel between x i and y j, and b j is the bi as. The symbol $*$ indicates convolutional operation. When there are M input maps and N output maps, this layer will contain N 3D kernels of size d ×d ×M (d ×d is the size of local receptive fields) and each kernel owns a bias. The last idea sub-sampling is also known as pooling, which can reduce data dimension. Max-pooling and average-pooling are popular operations in CNNs. As an example, the max-pooling operation is formulated as:

$$y^i_{r,c} = \max_{0 \le m,n < s} \{x^i_{r \cdot s+m, c \cdot s+n}\}, \tag{II.2}$$

where y i r,c is the neuron located at (r, c) in the i -th output map of a max-pooing layer, and it is assigned with the maximal value over a local region of size s ×s in the i -th input map x i . By combining the above three ideas, convolutional networks could obtain some important invariances on translation and scale to a certain degree. In [14], Krizhevsky et al. proposed a CNN model for image classification and achieved a landmark success. In the past three years, CNNs have been successfully introduced into various fields in computer vision from high-level tasks to low-level tasks, such as face detection [15], face recognition [16] , semantic segmentation [17] ,Super-resolution [18], patch similarity comparison [19], etc. These CNN-based methods usually outperform conventional methods in their respective fields, owning to the fast development of modern powerful GPUs, the great progress on effective training techniques, and the easy access to a large amount of image data. This study also benefits from these factors.

II.6. CNNs for image fusion

- **Feasibility**

As mentioned above, the generation of focus map in image fusion can be viewed as a classification problem [1]. Specifically, the activity level measurement is known as feature extraction, while the role of fusion rule is similar to that of a classifier used in general classification tasks. Thus, it is theoretically feasible to employ CNNs for image fusion. The CNN architecture for visual classification is an end-to-end framework [2], in which the input is an image while the output is a label vector that indicates the probability for each category. Between these two ends, the network consists of several convolutional layers (a non-linear layer like ReLU always follows a convolutional layer, so we don't explicitly mention it later), max-pooling layers and fully-connected layers. The convolutional and max-pooling layers are generally viewed as feature extraction part in the system, while the fully-connected layers existing at the output end are regarded as the classification part.

We further explain this point from the view of implementation. For most existing fusion methods, either in spatial domain or transform domain, the activity level measurement is essentially implemented by designing local filters to extract high-frequency details. On one hand, for most transform domain fusion methods, the images or image patches are represented using a set of pre- designed bases such as wavelet or trained dictionary atoms. From the view of image processing, this is generally equivalent to convolving them with those bases [18]. For example, the implementation of discrete wavelet transform is exactly based on filtering. On the other hand, for spatial domain fusion methods, the situation is even

clearer that so many activity level measurements are based on high-pass spatial filtering. Furthermore, the fusion rule, which is usually interpreted as the weight assignment strategy for different source images based on the calculated activity level measures, can be transformed into a filtering-based form as well. Considering that the basic operation in a CNN model is convolution (the full connection operation can be viewed as convolution with the kernel size that equals to the spatial size of input data [17]), it is practically feasible to apply CNNs to image fusion.

- **Superiority**

Similar to the situation in visual object classification applications, the advantages of CNN-based fusion method over existing methods are twofold. First, it overcomes the difficulty on manually designing complicated activity level measurement and fusion rules. The main task is replaced by the design of network architecture. With theemergence of some easy-to-use CNN platforms such as Caffe [21] and MatConvNet [20] , the implementation of network design becomes convenient to researchers. Second, and more importantly, the activity level measurement and fusion rule can be jointly generated via learning a CNN model. The learned result can be viewed as an "optimal" solution to some extent, and therefore is likely to be more effective than manually designed ones. Thus, the CNN-based method has a great potential to produce fusion results in higher quality than conventional methods.

II.7. Detailed fusion scheme

- **Focus detection**

Let A and B denote the two source images. In the proposed fusion algorithm, the source images are converted to grayscale space if they are color images. Let Aˆ and Bˆ denote the grayscale version of A and B (keep Aˆ = A and Bˆ = B when the source images are originally in grayscale space), respectively. A score map S is obtained by feeding Aˆ and Bˆ to the trained CNN model. The value of each coefficient in S ranges from 0 to 1, which suggests the focus property

Fig (II.10): Initial segmentation. (a) Focus map (b) binary segmentation map.

Of a pair of patches of size 16 ×16 in source images The closer the value is to 1 or 0, the more focused the patch from source image Aˆ or Bˆ is. For two neigh boring coefficients in S, their corresponding patches in each source image are overlapped with a stride of two pixels. To generate a focus map (denoted as M) with the same size of source images, we assign the value of each coefficient in S to all the pixels within its corresponding patch in M and average the overlapping pixels. **Fig (II.10)**(a) It can be seen that the focus information is accurately detected. Intuitively, the values of the regions with abundant details seems to be close to 1 (white) or 0 (black), while the plain regions tend to own values close to 0.5 (gray).

- **Initial segmentation**

To preserve useful information as much as possible, the focus map M needs to be further processed. In our method, as with most spatial domain multi-focus image fusion methods [22, 23, 24], we also adopt the popular "choose-max" strategy to pro- cess M. Accordingly, a fixed threshold of 0.5 is applied to segment M is into a binary map T, which is in accord with the classification principle of the learned CNN model. That is, the focus map is segmented by

$$T(x,y) = \begin{cases} 1, & M(x,y) > 0.5 \\ 0, & otherwise \end{cases} \quad \textbf{(II.3)}$$

The obtained binary map is shown in **Fig (II.10)**(b) (please notice the optical illusion in the focus map shown in **Fig (II.10)**(a), namely, the gray regions seems to be darker than its real intensity in a white background while brighter than its real intensity in a black back- ground). It can be seen that almost all the gray pixels in the focus map are correctly classified, which demonstrates that the learned CNN model can obtain precise performance even for the plain regions in source images.

- **Consistency verification**

It can be seen from **Fig (II.10)**(b) that the binary segmented map is likely to contain some misclassified pixels, which can be easily removed using the small region removal strategy.

Fig (II.11): Consistency verification and fusion (a) Initial decision map (b) Initial fused image (c) final decision map (d) fused image.

Specifically, a region which is smaller than an area threshold is reversed in the binary map. One may notice that the source images sometimes happen to contain very small holes. When this rare situation occurs, users can manually adjust the threshold even to zero, which means the region removal strategy is not applied. We will show in the next Section that the binary classification results can already achieve high accuracy. In this paper, the area threshold is universally set to **0.01 ×H ×W**, where **H** and **W** are the height and width of each source image, respectively. **Fig (II.11)**(a) shows the obtained initial decision map after applying this strategy. **Fig (II.11)**(b) shows the fused image using the initial decision map with the weighted-average rule. It can be seen that there are some undesirable artefacts around the boundaries between focused and defocused regions. Similar to [22], we also take advantage of the guided filter to improve the quality of initial decision map. Guided filter is a very efficient edge-preserving filter, which can transfer the structural information of a guidance image into the filtering result of the input image. The initial fused image is employed as the guidance image to guide the filtering of initial decision map. There are two free parameters in the guided filtering algorithm: the local window radius **r**and the regularization parameter**ε**. In this work, we experimentally set **r** to 8 and **ε** to 0.1. **Fig (II.11)**(c) shows the filtering result of the initial decision map given in **Fig (II.11)**(b).

- **Fusion**

Finally, with the obtained decision map D, we calculate the fused image F with the following pixel-wise weighted-average rule

$$\mathbf{F(x.y) = D(x.y)\ A(x.y) + (1- D(x.y))\ B(\ x.y)} \qquad \mathbf{(II.4)}$$

The fused image of the given example is shown in **Fig (II.11)**(d).

II.8. Conclusion

Convolutional Neural Networks(CNN) has become state-of-the-art algorithm for computer vision, natural language processing, and pattern recognition problems. This CNN has been using to build many use cases models from simply digit recognition to complex medical image analysis. This chapter tried to explain each components of a CNN, how it works to image analysis, and other relevant things. This chapter also gives a review from foundation of CNN to latest models and mentioned some applications areas

Chapter 3: Result and discussion

III.1 Introduction

In this chapter , will present the simulation results of method CNN which it combines the advantages of several metrics to provide better quality image. Next , will test our extension of the CNN method for fusing images.

III.2 Experiment and Dataset

In this section, we provide a study for multi-focus image fusion.The performance of CNN methods on 5 commonly-used multi-focus imagesare evaluated using 3 popular objective metrics. In our experiments, 5 pairs of multi-focus images that are widely used in the literatureof multi-focus image fusion are employed, which have been adopted in this field for years.Such as the "clock" pair, the "rose" pair, the "book" pair. The database reference [25].

III.3 Proposed method

Multi-focus image fusion has attracted increasing interests in image processing andcomputer vision. This chapter propose a multi-focus image fusion method based on focusconvolutional neural network (CNN). So we have to determine the fusion of the images to get a more clean and precise images using the CNN method. The important steps of this method are summarized as follows:

Firstly, Focus detection is to determine whether an image is in focus or not. Focus detection is able to be used for improving camera autofocus performance.

Secondly, an Initial segmentation is obtained by taking the pixel-wise maximum rule of the corresponding refined focus map.

Thirdly, Consistency verification is optimized into a final using region removal strategy and guided filter.

Finally, fusion imageprocess is defined as gathering all the important information from multiple images and their inclusion into fewer images

III.4 parameters setting

The proposed fusion method has two key parameters to be set the local window radius **r** and regularization factor **ε** ofkernel function for guide filter. During experiment, we use 5 pairs multi-focus images as test images, and calculate the above three objective evaluation metrics (view chapter1)to evaluate the influence of two parameters on fusion performance. **r**isset to **8** and **ε** is set to **0.1**.

III.5 Experimental settings

In order to evaluate the performance of the proposed method, 5 pairs of public multifocusimage used as test images are shown in **fig (III.1).** CNN is based on focusregion detection, and CNN is the most recently proposed method.To obtain the results of this methodwe use the original codes of this method,which are provided online by their authors.

(1) Book **(2) Calendar**

(3) Clock **(4) Rose**

(5) Blimp

Fig (III.1): 5 pairs of multi-focus image used as test images.

III.6Method validation

The assessment of the fused image quality is based on visual analysis and quantitativeanalysis. In quantitative analysis we used several statistical metrics **(QAB/F, LAB/F, NAB/F)**.In qualitative analysis, it depends on visuality to locate and identify faults that may affectthe image quality. This analysis is necessary to verify thequality of the images obtained by the fusion.

III.7 Quantitative analysis

The objective performance evaluation is included in this subsection, like shown in **table (III.2),** list of the values of three objective evaluation metrics **(QAB/F, LAB/F, NAB/F)** with method CNN, on 5 pairs multi-focus images. Meanwhile, **Table(III.1)** lists of the average values of the three metrics. The proposed fusion method CNN performs well in both subjective and objective evaluation, and present satisfactory fusion effects.

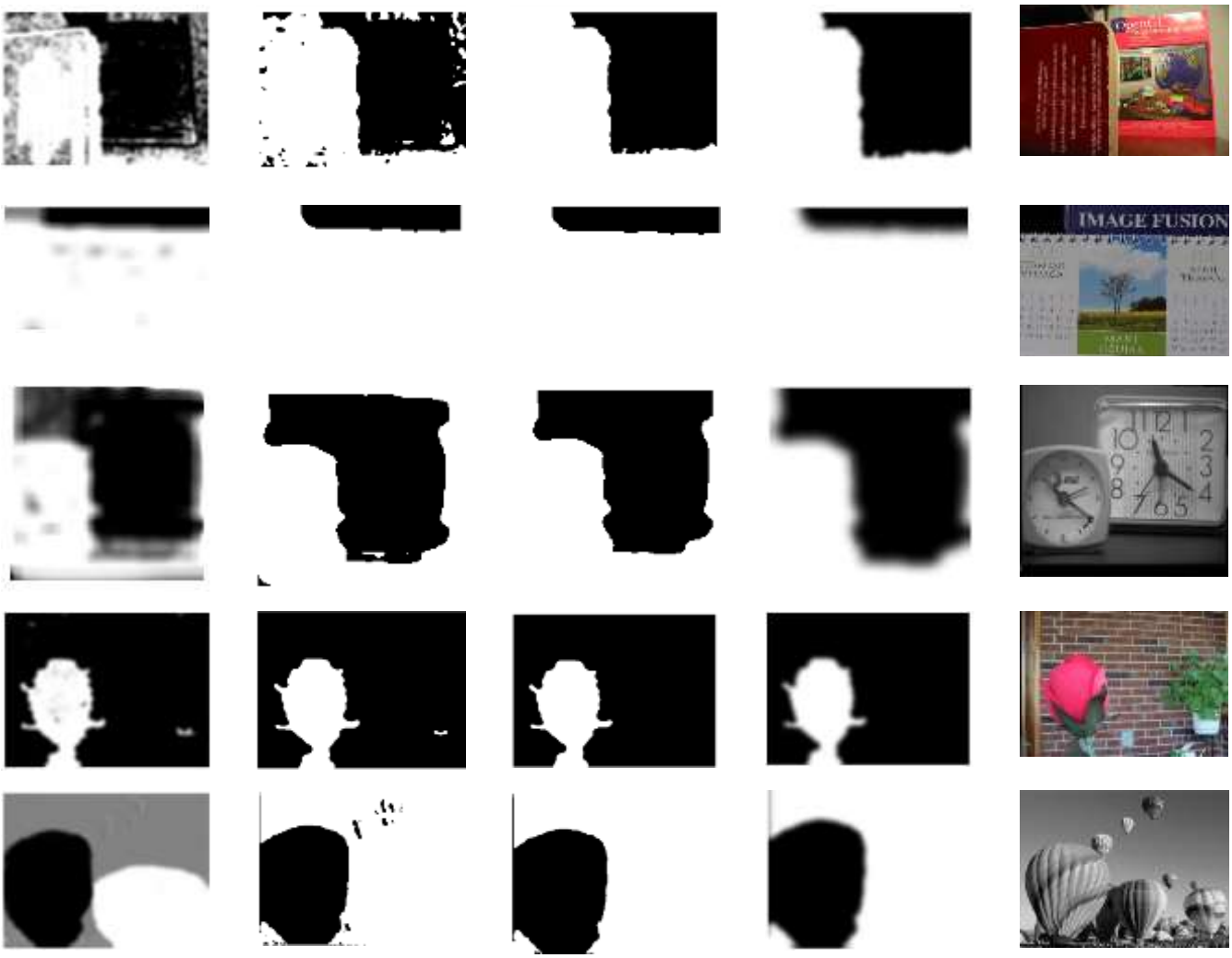
Metrics Pairs	QAB/F	LAB/F	NAB/F
CNN	0.8897	0.1103	0.0000

Table (III.1): Average evaluation of the three metrics on 5 pairs color images

Metrics Pairs	Q/AB/F	LAB/F	NAB/F
1	0.8754	0.1246	0.0000
2	0.8765	0.1233	0.0001
3	0.9026	0.0973	0.0001
4	0.8667	0.1333	0.0000
5	0.9272	0.0728	0.0000

Table (III.2): Evaluation metrics of CNN methods on 5 pairs color image

III.8 Intermediate results of the proposed method



Focus map | Binary segmented map | Initial decision map | Final decision map | Fused image

Fig (III.2): Intermediate fusion results of the proposed method.

To further exhibit the effectiveness of the CNN model for multi- focus image fusion, the intermediate results of five pairs of source images are given in **Fig (III.1)**

Therefore, for multi-focus image fusion, the binary segmented map can be interpreted as actual output of our CNN model. It can be seen from the second column of **Fig (III.2)** that the obtained segmented maps are very accurate that most pixels arecorrectly classified, which demonstrates the good capacity of the learned CNN model.

Nevertheless, there still exist two defects in terms of the binary segmented maps. First, a few of pixels are sometimes mi classified, leading to the emergence of small regions or holes in the segmented maps. These pixels usually locate at the plain regions of the scene that the distinction between two source images are very small. Among the five pairs of images, such situation arises in four of them except for the "**Calendar**"set. Since those misclassified pixels take up a very small proportion and they are usually locate at the plain regions, their impact on the fusion result is actually very slight. Even so, we apply the small region removal strategy to rectify these pixels. The third column in **Fig (III.2)**shows obtained initial decision map after this correction. Second, the boundaries between focused and defocused regions usually suffer from slight blocking artefacts. . Compared with the first defect, this one is more urgent to be addressed as the fusion quality around the boundaries is more important. Fortunately, edge-preserving filtering offers an appropriate tool to solve this problem.

The final decision maps shown in the fourth column of**Fig (III.2)**obtained with the guided filter are more natural in the boundary regions, leading to high visual quality of the fused results shown in the last column of **Fig (III.2)**.

III.9 Qualitative analysis

The fused results of fusion method for these five examples are shown in **fig (III.3).** Show two source multi-focus images for these five examples are shown in **fig (III.1)** we note clarity and consistency in the results. For example, in the pair (2) calendar, seen that the numbers are not clear, and the second picture, seen that the writing is not clear, as shown in **fig (III.1),** as well as for the other 4 pairs, but after fused image with CNN method, seen the clarity and consistency of the image as shown in **fig (III.3)**

(a) Book **(b) Calendar**

(c) Clock **(d) Rose**

(e) Blimp

Fig (III.3): resultants of 5 pairs of multi-focus image used as test images.

- **Metrics result**

Fig (III.4), (III.5), (III.6). Provide more insights about the objective performanceof CNN fusion method dataset. For each metric, the values obtainedby a method on different source images are connected to generate a curve and the averagevalue is given in the legend.

Almost CNN fusion method can obtain stable performance over all the testing images while only very few exceptions.

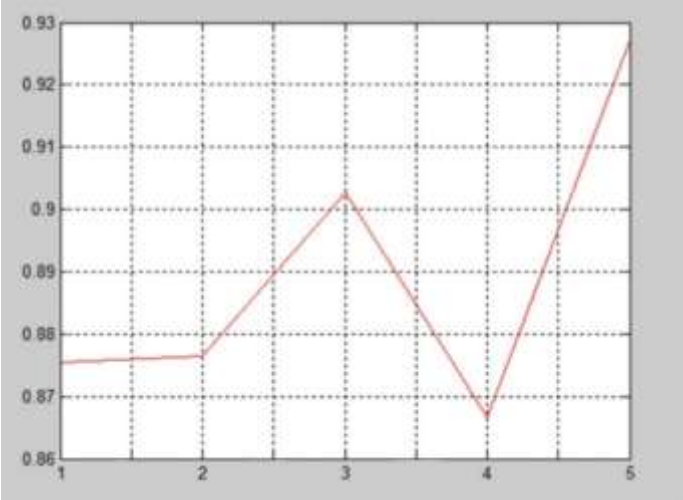


Fig (III.4): Objective performance of different fusion methods on metricTotal fusion performanceQAB/F

In **Fig (III.4)**, it can be noticed a small change in the resultmetric Total fusion performanceQAB/F between the 5 pairs with convergence in the results in the (1) Book. (2) Calendar. (4) Roseand also in the (3) Clock. (5) Blimp.

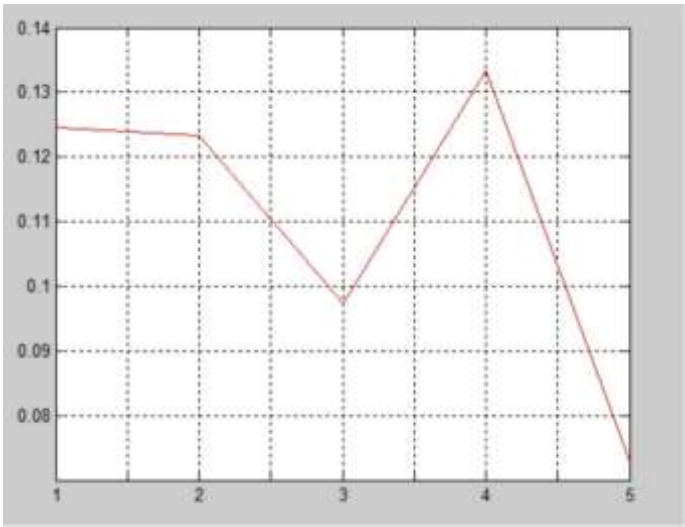


Fig (III.5): Objective performance of different fusion methods on metricLoss of fusion LAB/F

In **Fig (III.5)**,it can be noticed a small change in the resultmetric Loss of fusionLAB/F between the 5 pairs with convergence in the results in the (1) Book. (2) Calendar. (4) Rose and also in the (3) Clock. (5) Blimp

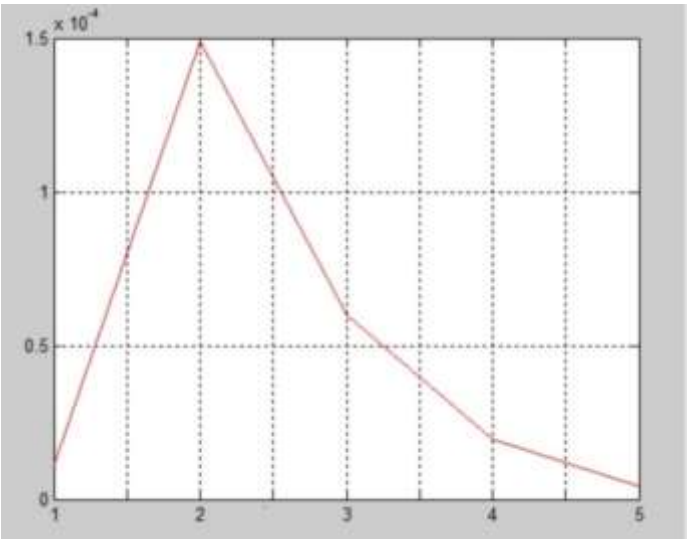


Fig (III.6): Objective performance of different fusion methods on metricArtefacts de la fusion$\boldsymbol{N}$AB/F.

In **Fig (III.6)** it can be noticed on a very small change in calendar (2) and clock (3) and the results in the book (1) and rose (4) and Blimp (5) we did not seen any change Value 0.

III.10 Conclusion

Image fusion is a process where we can fuse two or more images source coming fromdifferent or same image sensors so as to get a new image which contains more informationthan any of the original. In this last chapter, which consists of the implementationproposed method CNN on multi-focus, we have achieved the best results.

Final conclusion

In This work mainly presents a new multi-focus image fusion method based on a deep convolutional neural network. The main novelty of our method is learning a CNN model to achieve a direct mapping between source images and the focus map. Based on this idea, the activity level measurement and fusion rule can be jointly generated by learning the CNN model, which can overcome the difficulty faced by the existing fusion methods. The main contribution of this work could be summarized into the following four points:

1) Introduce CNNs into the field of image fusion. The feasibility and superiority of CNNs used for image fusion are discussed. It is the first time that CNNs are employed for an image fusion task to the best of our knowledge.

2) Propose a multi-focus image fusion method based on a CNN model. Experimental results demonstrate the proposed method can achieve state-of-the-art results in terms of visual quality and objective assessment.

3) Exhibit the potential of the learned CNN model for other-type image fusion is- sues.

4) Put forward some suggestions on the future study of CNN-based image fusion.

References

[1] Y. LeCun , L. Bottou , Y. Bengio , P. Haffner ,Gradient-based leaning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324 .

[2] S. Li , J. Kwok , Y. Wang , Multifocus image fusion using artificial neural net- works, Pattern Recognit. Lett. 23 (8) (2002) 985–997 .

[3] Zheng, Yufeng; Blasch, Erik; Liu, Zheng (2018). Multispectral Image Fusion and Colorization. SPIE Press. ISBN 9781510619067.

[4] Haghighat, M. B. A.; Aghagolzadeh, A.; Seyedarabi, H. (2011). "Multi-focus image fusion for visual sensor networks in DCT domain". Computers & Electrical Engineering. 37 (5): 789–797. doi:10.1016/j.compeleceng.2011.04.016.

[5] Haghighat, M. B. A.; Aghagolzadeh, A.; Seyedarabi, H. (2011). "A non-reference image fusion metric based on mutual information of image features" . Computers & Electrical Engineering. 37 (5): 744 756 doi:10.1016/j.compeleceng.2011.07.012

[6] M., Amin-Naji; A., Aghagolzadeh (2018). "Multi-Focus Image Fusion in DCT Domain using Variance and Energy of Laplacian and Correlation Coefficient for Visual Sensor Networks". Journal of AI and Data Mining. 6 (2): 233–250. doi:10.22044/jadm.2017.5169.1624. ISSN 2322-5211.

[7] Jasiunas MD, Kearney DA, Hopf J, Wigley GB (2002) Imagefusion for uninhabited airborne vehicles. In: 2002 IEEE Internationalconference on field-programmable technology, 2002.(FPT). Proceedings, p 348–351. IEEE

[8] Dong J, Dafang Z, Yaohuan H, Jinying F (2011) Survey of multispectral image fusion techniques in remote sensing applications. In: Zheng Y (ed) Image fusion and its applications. Alcorn State University, USA

[9] . Banu RS (2011) Medical image fusion by the analysis of pixel level multi-sensor using discrete wavelet Transform. In: Proceedings of the national conference on emerging trends in computing science, p 291–297

[10] Bavachan B, Krishnan DP (2014) A survey on image fusion techniques. IJRCCT 3(3):049–052

[11] A. SOUALAH , Chahinez, DERDACH, “Méthodes de fusion d ’ images basée sur la détection de saillance visuelle,”universitè ouargla 2019.

[12] F. Sultana, A. Su an, and P. Dutta Advancements in image classification using convolutional neural network. In 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pages 122{129, Nov 2018.

[13] V. Nair , G. Hinton , Rectified linear units improve restricted boltzmann ma- chines, in: Proceedings of 27th International Conference on Machine Learning, 2010, pp. 807–814 .

[14] A. Krizhevsky , I. Sutskever , G. Hinton , Imagenet classification with deep convo- lutional neural networks, in: Advances in Neural Information Processing Sys- tems, 2012, pp. 1097–1105 .

[15] S. Farfade , M. Saberian , L. Li , Multi-view face detection using deep convolu- tional neural networks, in: Proceedings of the 5th ACM on International Con- ference on Multimedia Retrieval, 2015, pp. 643–650 .

[16] Y. Sun , X. Wang , X. Tang , Deep learning face representation from predicting 10,0 0 0 classes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.

[17] J. Long , E. Shelhamer , T. Darrell , Fully convolutional networks for semantic seg- mentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[18] C. Dong , C. Loy , K. He , X. Tang , Image super-resolution using deep convolu- tional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307 .

[19] S. Zagoruyko , N. Komodakis , Learning to compare image patches via convolu- tional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4353–4361 .

[20] http://www.vlfeat.org/matconvnet/ .

[21] M. Nejati , S. Samavi , S. Shirani , Multi-focus image fusion using dictio- nary-based sparse representation, Inf. Fusion 25 (1) (2015) 72–84 .

[22] S. Li , B. Yang , Multifocus image fusion using region segmentation and spatial frequency, Image Vis. Comput. 26 (7) (2008) 971–979 .

[23] Z. Zhou , S. Li , B. Wang , Multi-scale weighted gradient-based fusion for multi–focus images, Inf. Fusion 20 (1) (2014) 60–72 .

[24] Y. Liu , S. Liu , Z. Wang , Multi-focus image fusion with dense sift, Inf. Fusion 23 (1) (2015) 139–155 .

[25] https://www.mathworks.com/matlabcentral/fileexchange/70109-multi-focus-image-fusion-dataset.

A b s t r a c t

As is well known, activity level measurement and fusion rule are two crucial factors in image fusion. For most existing fusion methods, either in spatial domain or in a transform domain like wavelet, the activity level measurement is essentially implemented by designing local filters to extract high-frequency details, or the calculated clarity information of different source images are then compared using some elaborately designed rules to obtain a clarity/focus map. Consequently, the focus map contains the integrated clarity information, which is of great significance to various image fusion issues, such as multi-focus image fusion, multi-modal image fusion, etc. However, in order to achieve a satisfactory fusion performance, these two tasks are usually difficult to finish. In this study, we address this problem with a deep learning approach, aiming to learn a direct mapping between source images and focus map. To this end, a deep convolutional neural network (CNN) trained by high-quality image patches and their blurred versions is adopted to encode the mapping. The main novelty of this idea is that the activity level measurement and fusion rule can be jointly generated through learning a CNN model, which overcomes the difficulty faced by the existing fusion methods. Based on the above idea, a new multi-focus image fusion method is primarily proposed in this paper. Experimental results demonstrate that the proposed method can obtain state-of-the-art fusion performance in terms of both visual quality and objective assessment. The computational speed of the proposed method using parallel computing is fast enough for practical usage. The potential of the learned CNN model for some other-type image fusion issues is also briefly exhibited in the experiments.

الـمـلـخص

كما هو معروف جيدًا ، يعد قياس مستوى النشاط وقاعدة الاندماج عاملين حاسمين في دمج الصورة. بالنسبة لمعظم طرق الاندماج الحالية ، سواء في المجال المكاني أو في مجال تحويل مثل المويجة ، يتم تنفيذ قياس مستوى النشاط بشكل أساسي من خلال تصميم المرشحات المحلية لاستخراج تفاصيل عالية التردد ، أو تتم مقارنة معلومات الوضوح المحسوبة لصور المصدر المختلفة باستخدام بعض قواعد مصممة بشكل متقن للحصول على خريطة الوضوح / التركيز. وبالتالي ، تحتوي خريطة التركيز البؤري على معلومات الوضوح المتكاملة ، والتي لها أهمية كبيرة في العديد من مشكلات دمج الصور ، مثل دمج الصور متعددة البؤرة ، ودمج الصور متعدد الوسائط ، وما إلى ذلك ، ومع ذلك ، من أجل تحقيق أداء اندماج مرضٍ ، فإن هذه عادة ما يكون من الصعب إنهاء مهمتين. في هذه الدراسة ، نعالج هذه المشكلة من خلال نهج التعلم العميق ، بهدف تعلم رسم الخرائط المباشر بين الصور المصدر وخريطة التركيز. تحقيقا لهذه الغاية ، تم اعتماد شبكة عصبية تلافيفية عميقة (CNN) مدربة بواسطة تصحيحات صور عالية الجودة وإصداراتها غير الواضحة لتشفير التعيين. الحداثة الرئيسية لهذه الفكرة هي أنه

يمكن إنشاء قياس مستوى النشاط وقاعدة الاندماج بشكل مشترك من خلال تعلم نموذج CNN ، والذي يتغلب على الصعوبة التي تواجهها طرق الاندماج الحالية. بناءً على الفكرة أعلاه ، تم اقتراح طريقة جديدة لدمج الصور متعددة التركيز بشكل أساسي في هذه الورقة. تظهر النتائج التجريبية أن الطريقة المقترحة يمكن أن تحصل على أحدث أداء اندماج من حيث الجودة البصرية والتقييم الموضوعي. السرعة الحسابية للطريقة المقترحة باستخدام الحوسبة المتوازية سريعة بما يكفي للاستخدام العملي. يتم أيضًا عرض إمكانات نموذج CNN الذي تم تعلمه لبعض مشكلات دمج الصور من النوع الآخر لفترة وجيزة في التجارب.

Résumé

Comme on le sait, la mesure du niveau d'activité et la règle de fusion sont deux facteurs cruciaux dans la fusion d'images. Pour la plupart des méthodes de fusion existantes, que ce soit dans le domaine spatial ou dans un domaine de transformation comme l'ondelette, la mesure du niveau d'activité est essentiellement mise en œuvre en concevant des filtres locaux pour extraire des détails à haute fréquence, ou les informations de clarté calculées de différentes images sources sont ensuite comparées à l'aide de certains des règles élaborées pour obtenir une carte de clarté/focus. Par conséquent, la carte de mise au point contient les informations de clarté intégrées, ce qui est d'une grande importance pour divers problèmes de fusion d'images, tels que la fusion d'images multi-foyers, la fusion d'images multimodales, etc. Cependant, afin d'obtenir des performances de fusion satisfaisantes, ces deux tâches sont généralement difficiles à terminer. Dans cette étude, nous abordons ce problème avec une approche d'apprentissage en profondeur, visant à apprendre une correspondance directe entre les images sources et la carte de mise au point. À cette fin, un réseau de neurones à convolution profonde (CNN) formé par des patchs d'images de haute qualité et leurs versions floues est adopté pour coder la cartographie. La principale nouveauté de cette idée est que la mesure du niveau d'activité et la règle de fusion peuvent être générées conjointement par l'apprentissage d'un modèle CNN, ce qui surmonte la difficulté rencontrée par les méthodes de fusion existantes. Sur la base de l'idée ci-dessus, une nouvelle méthode de fusion d'images multi-foyers est principalement proposée dans cet article. Les résultats expérimentaux démontrent que la méthode proposée peut obtenir des performances de fusion de pointe en termes de qualité visuelle et d'évaluation objective. La vitesse de calcul de la méthode proposée utilisant le calcul parallèle est suffisamment rapide pour une utilisation pratique. Le potentiel du modèle CNN appris pour certains problèmes de fusion d'images d'un autre type est également brièvement exposé dans les expériences.