Information Fusion 42 (2018) 158–173

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Deep learning for pixel-level image fusion: Recent advances and future prospects

Yu Liu[a], Xun Chen[a,b,*], Zengfu Wang[c], Z. Jane Wang[d], Rabab K. Ward[d], Xuesong Wang[e,*]

[a] Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China
[b] Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230026, China
[c] Department of Automation, University of Science and Technology of China, Hefei 230026, China
[d] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
[e] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Keywords:
Image fusion
Deep learning
Convolutional neural network
Convolutional sparse representation
Stacked autoencoder

ABSTRACT

By integrating the information contained in multiple images of the same scene into one composite image, pixel-level image fusion is recognized as having high significance in a variety of fields including medical imaging, digital photography, remote sensing, video surveillance, etc. In recent years, deep learning (DL) has achieved great success in a number of computer vision and image processing problems. The application of DL techniques in the field of pixel-level image fusion has also emerged as an active topic in the last three years. This survey paper presents a systematic review of the DL-based pixel-level image fusion literature. Specifically, we first summarize the main difficulties that exist in conventional image fusion research and discuss the advantages that DL can offer to address each of these problems. Then, the recent achievements in DL-based image fusion are reviewed in detail. More than a dozen recently proposed image fusion methods based on DL techniques including convolutional neural networks (CNNs), convolutional sparse representation (CSR) and stacked autoencoders (SAEs) are introduced. At last, by summarizing the existing DL-based image fusion methods into several generic frameworks and presenting a potential DL-based framework for developing objective evaluation metrics, we put forward some prospects for the future study on this topic. The key issues and challenges that exist in each framework are discussed.

1. Introduction

The aim of pixel-level image fusion[1] is to generate a composite image from multiple input images containing complementary information of the same scene [1]. The input images known as source images are captured from different imaging devices or a single type of sensor under different parameter settings. The composite image known as fused image should be more suitable for human or machine perception than any individual input. Due to this advantage, image fusion techniques exhibit great significance in a variety of applications that rely on two or more images of the same scene. For instance, in clinical diagnosis, physicians usually need medical images obtained by different modalities including computed tomography (CT), magnetic resonance (MR), single photon emission computed tomography (SPECT), etc. In this situation, integrating the important information of different source images into a composite image can often reduce the difficulty in achieving precise diagnosis. Another typical application of image fusion is digital photography, for example to extend the depth-of-field or dynamic range of a camera. Some other popular scenarios of image fusion include video surveillance, remote sensing, etc.

The study of image fusion has lasted for more than 30 years, during which hundreds of related scientific papers have been published [2]. In recent years, deep learning (DL) has gained many breakthroughs in various computer vision and image processing problems, such as classification [3], segmentation [4], super-resolution [5], etc. In the field of image fusion, the study based on deep learning has also become an active topic in the last three years. A variety of DL-based image fusion methods have been proposed for digital photography (e.g., multi-focus image fusion, multi-exposure image fusion) [6–9], multi-modality imaging (e.g., medical image fusion, infrared/visible image fusion) [10–12], and

* Corresponding authors at: Department of Biomedical Engineering, Hefei University of Technology, Hefei 230 0 09, China (X. Chen) and School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (X. Wang).
E-mail addresses: yuliu@hfut.edu.cn (Y. Liu), xunchen@ece.ubc.ca (X. Chen), zfwang@ustc.edu.cn (Z. Wang), zjanew@ece.ubc.ca (Z.J. Wang), rababw@ece.ubc.ca (R.K. Ward), wangxuesongcumt@163.com (X. Wang).

[1] It is known that image fusion can be grouped into three categories, namely, pixel-level fusion, feature-level fusion and decision-level fusion. In this paper, we focus on pixel-level image fusion. For simplicity, we omit the term pixel-level in most expressions later.

https://doi.org/10.1016/j.inffus.2017.10.007
Received 26 September 2017; Received in revised form 19 October 2017; Accepted 24 October 2017
1566-2535/

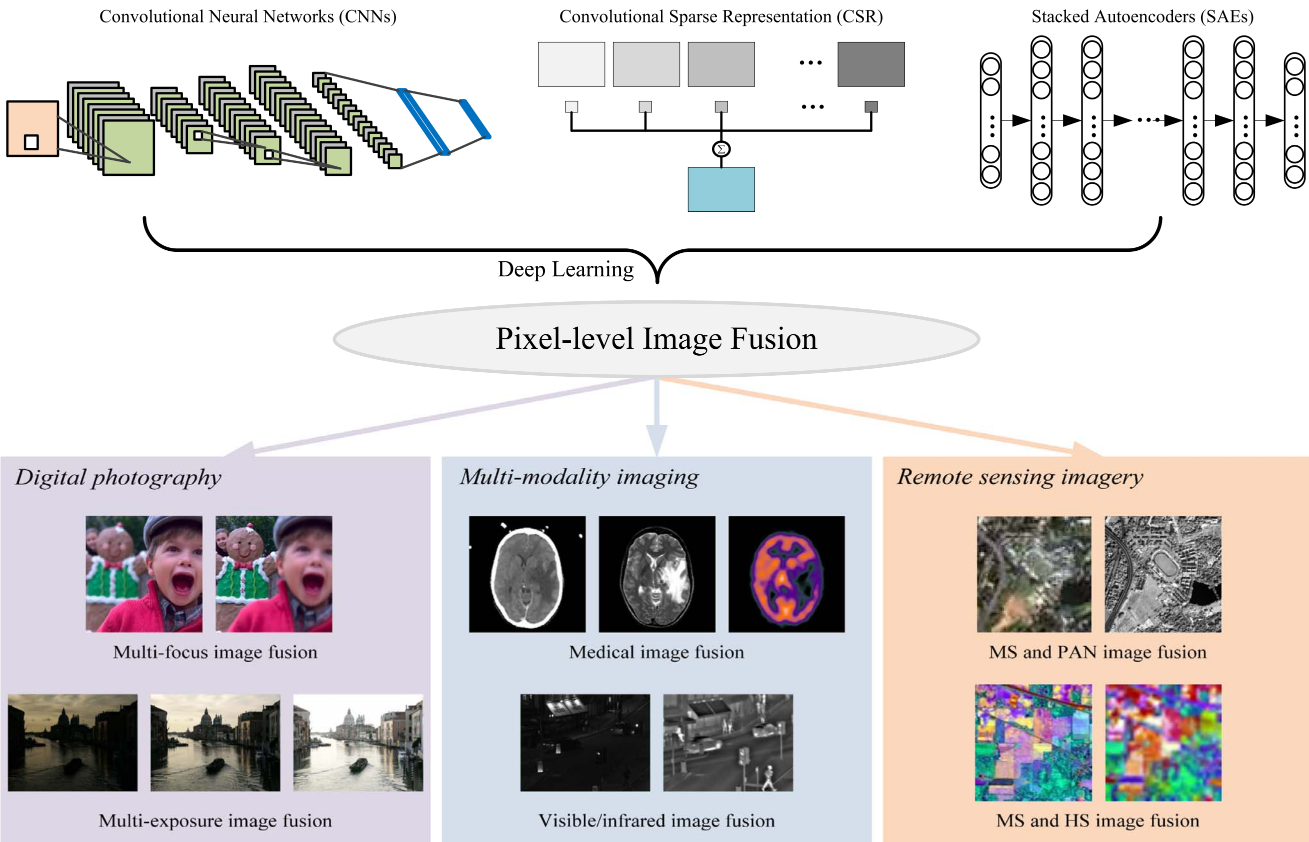

Fig. 1. An illustration of DL-based image fusion.

remote sensing imagery (e.g., multi-spectral (MS) and panchromatic (PAN) image fusion, MS and hyper-spectral (HS) image fusion) [13–19], showing advantages over conventional methods and leading to state-of-the-art results. Fig. 1 shows an illustration of DL-based image fusion. In this paper, we will review the recent advances related to DL-based image fusion and put forward some future prospects on this topic.

A number of representative survey works concerning image fusion have been proposed in the literature. In 1999, Zhang and Blum performed a detailed review on multi-scale decomposition (MSD)-based image fusion approaches [20]. Another influential survey on multi-resolution image fusion was presented by Piella in [21]. Liu et al. conducted a thorough study of the objective metrics used in fusion performance evaluation [22]. Zhang et al. gave a comprehensive review on the sparse representation (SR)-based image fusion methods [23]. Li et al. recently presented an all-round survey about image fusion covering methods, objective metrics and applications [2]. There are also arising some surveys that concentrate on the fusion issues in some specific application fields, such as medical imaging [24,25], remote sensing [26,27] and surveillance [28]. As DL-based image fusion has just been studied very recently, the related methods are not included in existing surveys related to image fusion. This paper presents a specific review of the recent achievements in DL-based image fusion, aiming to give a comprehensive introduction about the current progress in this field. Furthermore, we put forward several specific prospects for the future study of DL-based image fusion, hoping to provide some new thoughts for researchers in the field of image fusion.

In addition to the popularity of DL-based approaches for image processing, another important factor that motivates the study on DL-based image fusion is overcoming the difficulties faced by conventional image fusion research[2]. Therefore, we first summarize the related difficulties and describe the Advantages of DL for image fusion accordingly in Section 2. In Section 3, some commonly-used deep learning models in image fusion study are briefly introduced. Section 4 presents a detailed review of the recent advances in DL-based image fusion. In Section 5, we put forward some prospects for the future study of DL-based image fusion. Finally, concluding remarks are given in Section 6.

2. Motivations of DL for image fusion

2.1. Difficulties in conventional image fusion research

The study of image fusion generally contains two aspects: image fusion methods and objective evaluation metrics. In this subsection, we discuss the main difficulties that exist in these two aspects respectively. For better understanding of the discussions, some representative methods will be briefly introduced, while more comprehensive and more detailed reviews could be found in the survey papers mentioned above.

2.1.1. Image fusion methods

In view of its different application fields, image fusion can be categorized into different sub-problems, such as multi-focus image fusion, multi-exposure image fusion, medical image fusion, visible/infrared image fusion, remote sensing image fusion, etc. Among them, remote sensing image fusion, often known as pan-sharpening in most cases, has distinct differences with the other types of fusion problems. Pan-sharpening indicates the process of fusing a low-resolution MS image and a high-resolution PAN image to obtain a MS image with high spatial resolution. Thus, pan-sharpening can be viewed as a super-resolution problem for the MS image aided by a PAN image, and many methods originating from the field of image super-resolution have been applied to remote sensing image fusion. The situation of MS and HS image fusion is quiet similar, which can be regarded as a super-resolution

[2] The term *conventional image fusion research* in this paper indicates the research that is not based on deep learning approaches.