Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.54
https://jatuc.uoturath.edu.iq/

An Evaluation of Developments in Image Processing Through Artificial Intelligence Methods: Review

Amal Abbas Kadhim[1*], Wedad Abdul Khuder Naser[2], Safana Hyder Abbas[3]

[1,2,3]Department of Computer Science, College of Education, Mustansiriyah University, Baghdad, Iraq.

[*]*Correspondence email*: abbas.amal@uomustansiriyah.edu.iq

KEYWORDS

Artificial Intelligence; Image Processing; Machine Learning; Deep Learning; Convolutional Neural Networks; Transformer Models.

ABSTRACT

Artificial intelligence has been recognized as a major force in revolutionizing image processing. Hence, this paper undertakes a survey of recent image processing technologies, which mainly include Artificial Intelligence techniques like machine learning and deep learning methods. Recent AI-based technologies like Convolutional Neural Networks (CNNs) and transformer-based models have significantly raised the level of performing different image tasks such as image classification, object detection, image enhancement, and pattern recognition. In addition to the High-tech in Medical Imaging Security autonomous systems, and multimedia technologies are also regarded as the main application domains of AI-driven image processing in this review. Highlight the main issues such as very large dataset dependence, high computational cost and incomprehensibility of models. At last, we are talking about the good ways of future research development that aim at the promotion of able, scalable and interpretable AI models. In brief, this article gives a full review of AI as one of the most effective change agents in the image processing field.

1. INTRODUCTION

Artificial intelligence (AI) is one of the most potent technologies at our disposal, totally transforming the domain of computing, including image processing. In simple terms, image processing is basically changing and analyzing digital pictures so that information can be conveyed visually. In fact, it is an indispensable part in a wide range of areas, including medical diagnosis surveillance, as well as in driverless or self-driving cars. At the same time, with the rapid development of AI algorithms, conventional methods for image processing have been transformed dramatically and in a lot of cases totally replaced by data-driven techniques [1].

Machine learning and deep learning algorithms, mainly CNNs, have done very well on complex visual tasks. For instance, CNNs have been highly efficient in learning hierarchical image features directly from the raw data without any manual effort [2]. Therefore, these technological evolutions have played a major role in improving image classification, object detection, and segmentation, etc. significantly.

Transformer-based architectures have pretty clearly turned out to be the main choice over CNNs, mainly when one is looking at a global context and also uses them in computer vision tasks. This is just because these models can capture long-range dependencies within images and, as a result, are very capable of producing state-of-the-art results [3]. Artificial intelligence in image processing has found its major applications in areas like medical imaging for the detection of diseases at a very early stage. Besides, security and autonomous systems depend on photo analysis that is done instantly [4]. However, the challenges of extremely expensive computer operations, the requirement for a large amount of data, and low interpretability still remain.

1.1 Research Contribution

The paper presents a detailed and exhaustive survey of the literature on image processing using Artificial Intelligence. One of the ways the review is performed is by separating and categorizing the main methods (such as CNNs, GANs, Transformers, Transfer Learning), determining their strengths and weaknesses, identifying the areas where more research is required, and highlighting the trends of AI hybrid systems.

2. AI SUBSISTENCE ARCHITECTURES FOR IMAGE PROCESSING

In recent times, the application of artificial intelligence techniques has led to major breakthroughs in image processing, particularly with the launch of deep learning models. A major milestone was the 2012 paper by Krizhevsky et al. on AlexNet. The architecture of AlexNet demonstrated the remarkable ability of CNNs to handle very large image classification problems. Their method produced results that were much better than those of conventional image processing methods, and it is considered the dawn of a new period in computer vision [5].

After this, Goodfellow et al. (2014) came up with Generative Adversarial Networks (GANs), a new way of creating very realistic images by having two neural networks train each other through competition. Since then, this model has been extensively used in image enhancement, reconstruction, and synthesis, achieving amazing results in visual quality [6].

Besides, Simonyan and Zisserman (2015) created the VGGNet model that highlighted the role of the network depth in enhancing the accuracy of image recognition. Their work showed that deeper CNN architectures could yield better scores but would come with significantly higher computational costs. In order to alleviate the issues of deep networks, He et al. (2016) introduced Residual Networks (ResNet), which not only tackled the degradation problem but also allowed for training very deep neural networks, thus achieving even better image classification results [7].
Besides, Long et al. (2015) introduced Fully Convolutional Networks (FCN) for semantic image segmentation, which allows the classification of pixels in an image, which leads the segmentation accuracy being highly improved. Likewise, Ronneberger et al. (2015) came up with a U-Net architecture. This turned out to be very powerful in biomedical image segmentation. Also, it was particularly good when there was limited training data [8].

Besides, Pan and Yang (2010) discussed Transfer Learning, emphasizing how it helps a lot in reusing pre-trained models for new image processing missions. It is also very practical for real-world applications because it lessens both the demand for massive datasets and the time for training [9].

Recently, Dosovitskiy et al. (2021) came up with the Vision Transformer (ViT), which is the adaptation of transformer-based architectures for the image processing domain. They revealed in their work that transformers could potentially match and even surpass the performance of CNN-based methods, thus opening the field to a novel direction [3].

Highly summarizing, artificial intelligence has radically changed the processing of images, which was done by a classical rule-based approach before that was changed to a data-driven approach by these researchers. However, disadvantages like high computational burden, data dependency, and model interpretability still are the main fields for research in the future, even after these changes [2][10].

3. IMAGE RESTORATION AND ENHANCEMENT

Image restoration and enhancement are among the critical functions of digital image processing intended to boost the image quality and retrieve damaged visual details. Image restoration concentrates on recovering the original image from its noisy or degraded version, which is usually caused by the addition of noise through some sort of image acquisition or transmission process or alteration of an image in some way. The degradation modes that are typically considered are Gaussian noise, motion blur, and defocus blur, which can all make the image less clear and less usable. [1]

Traditional methods of image restoration usually involve mathematical modeling and optimization techniques. Methods like inverse filtering and Wiener filtering are typical examples of techniques used to deblur and denoise images. Nevertheless, they struggle when the blurring model is either unknown or very complex [1]. Contrarily, image enhancement methods mainly aim at visually improving the image quality by removing the need for a physical degradation model. Most widely used techniques for increasing brightness, contrast, and enhancing edge details are histogram equalization, contrast stretching, and sharpening filters [11].

Actually, deep learning has revolutionized AI-based image restoration and enhancement techniques. By far, Convolutional Neural Networks (CNNs) are the top favorite tool that is utilized for various image processing tasks such as noise reduction, resolution enhancement, and deblurring. For example, both DnCNN and SRCNN have beaten conventional methods by learning complex transformations from degraded images to high-quality ones [12].

For instance, some of the latest techniques utilizing Generative Adversarial Networks (GANs) and transformer-based frameworks have been achieving top-notch results in the realm of image enhancement. To illustrate, GAN-driven models like SRGAN are capable of producing high-resolution images with lifelike textures, while transformers excel at identifying global features and grasping contexts [13], [14].

However, even with these improvements, there are still problems such as the high computational demands, the reliance on very large data sets, and the capability to deal with deteriorations in the real world. Therefore, current research emphasizes building models that not only are efficient and robust but can also be made understandable for medical imaging, surveillance, and multimedia systems applications [11], [13].

3.1 Super-Resolution

One of the most dynamic areas in image processing is super-resolution (SR), which aims at the generation of a high-resolution (HR) image that corresponds to a low-resolution (LR) version of the image. The development of this technique has been a huge asset to the medical field for visualization and has provided a great source of satellite imaging, monitoring, and multimedia systems. While the primary aim of the process is to enhance the spatial resolution, it is also a fine art to preserve the small details and textures that are usually the first to go when the images are captured or compressed [11].

Initially, super-resolution techniques were based on interpolation methods like bilinear and bicubic interpolation. But these methods usually result in blurry images and are not capable of regenerating the finest image details. The development of deep learning has led to the creation of sophisticated models that can learn intricate transformations from LR to HR images [1].

The Super-Resolution Convolutional Neural Network (SRCNN) is among the first deep learning models for super-resolution. It was Dong et al who introduced the idea of SRCNN as a model that can directly learn an end-to-end mapping between low- and high-resolution images using a fairly simple convolutional architecture of three layers. In spite of its simplicity, SRCNN notably surpasses conventional approaches with regard to both the quality of reconstruction and the time required [15].

By expanding upon CNN-based methods, powerful architectures like the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) have been put forward. ESRGAN takes a step further compared to the previous GAN-based models by adding residual-in-residual dense blocks and a relativistic discriminator, thus it can create more beautiful and more true-to-life details. The model is very good at enhancing the quality of the perception of images, so it is a great choice for scenarios where the look needs to be as real as possible [16].

Recently, models relying on Transformer architecture, such as SwinIR, have been leading image super-resolution tasks. SwinIR inherits the Swin Transformer architecture that relies on shifted window self-attention to effectively capture local as well as global context information. This progress in the model's capacity to generate fine and intricate structures with high accuracy reached beyond what CNN-based methods were capable of achieving [17].

In addition to these advancements, issues such as computational complexity, memory requirements, and the ability to handle unseen real-world degradations still remain the main research focus in this area. Therefore, more and more attention is being paid to the development of lightweight and efficient models that do not sacrifice the quality of image reconstruction [1], [17].

3.2 Noise Removal

Image denoising is one of the main preprocessing steps in image processing to get a clean image from a noisy one. Noise can be introduced to the image from various sources, such as imperfections of the sensor, low light conditions, and errors during transmission. The most common types of noise are Gaussian noise, salt-and-pepper noise, and Poisson noise [11].

Besides degrading the quality of the image, the noise may also impair other tasks such as segmentation and classification. The conventional denoising methods are basically the spatial filtering techniques, such as Gaussian filtering, median filtering, and non-local means. Methods like these undoubtedly could remove noise, but often they end up getting rid of the small details and textures [1]. Research into denoising has primarily been centered on classical methods such as non-local means (NLM) and BM3D, which are recognized as the state-of-the-art methods for moderately noisy images. Nonetheless, deep learning has rapidly become the dominant force in this area, bringing with it new models such as the Denoising Convolutional Neural Network (DnCNN) that have demonstrated extraordinary abilities. Rather than first generating a clean image and then removing the noise, DnCNN uses residual learning to directly estimate the noise. Besides enabling a quicker training process, residual learning also leads to better denoising results. Moreover, it hasn't been difficult for it to transfer its great denoising capacity to a variety of save noise levels and types [12]. Moreover, U-Net and its derivates are very popular for image denoising in the medical field, especially. The encoder-decoder architecture, combined with skip connections allow U-Net to capture not only

the overall context but also the fine details at a pixel level. This feature makes it very efficient in preserving structural elements while removing noise [18].

A model's capability to process blind denoising without the prior knowledge of noise is the next significant step in blind denoising. Deep learning-based blind denoising methods can handle various noise types by training the model on extensive and diverse data. They increase the robustness of the model and can be utilized in real scenarios where noise characteristics are unknown [17].

However, one of the main focuses of research still remains the issues of over-smoothing, loss of details, and computational complexity despite these enhancements. Nowadays, researchers are working on developing lightweight and flexible models that are suitable for real-time and resource-limited applications [1], [17].

3.3 De-blurring

Picture de-blurring is the process of restoring a crisp image from a blurred one. The blur can result from either camera shake, the object moving, or the image being out of focus. Motion blur and defocus blur are the two main types. They can be viewed as convolution operations between the underlying sharp image and the blur kernel [11].

Traditionally, de-blurring methods were largely based on deconvolution approaches such as Wiener filtering and the Richardson-Lucy algorithm. These techniques rely on having a very accurate estimate of the blur kernel, which is a tough task in real-life situations, and therefore, the restoration results are often not the best ones [1]. The power of deep learning has significantly upgraded the capabilities of image de-blurring. Convolutional Neural Networks (CNNs) are getting very popular for learning the direct mapping from a blurred image to a sharp image, thus eliminating the need for an explicit blur kernel estimation. Model taking DeblurGAN as the use of Generative Adversarial Networks (GANs) to produce images that are highly realistic and distinctly clear even under difficult circumstances [19]. In addition, the effectiveness of de-blurring is significantly improved by multi-scale and attention-based architectures which excel at locating both local and global image features. Besides, transformer-based models are top performers by being able to better manipulate spatial dependencies and the complicated character of blur, hence leading to even greater reconstruction fidelity [17].

While image de-blurring has seen considerable advancement, it remains a difficult challenge due to the inherent ill-posed nature of the problem, especially when the blur is, on the one hand, extremely severe and when on the other hand, non-uniform. Scientists are always seeking ways to develop much more robust techniques, with a smaller computational footprint and great at handling real-world blurred images, without closure.

3.4 Image Inpainting

Image inpainting is a basic function of image processing used for re-creating the disappearance or damage of image parts in a manner visually acceptable manner. It is about the replacement of the lost elements to make the content of the restored parts in harmony with the neighbours in relation to the texture, structure, and meaning. Such work is used for modifying images, erasing objects, bringing back old photographs, and covering the blind spots in computer vision systems [11].

Mostly, the traditional inpainting methods rely on diffusion and patch-based techniques. Diffusion-based approaches spread the pixel information from the surrounding regions to the missing ones; however, they often fail with quite large missing areas. Patch-based methods, conversely, fill the gaps by copying similar patches from the known regions, which are great for texture synthesis, but sometimes they are not good enough to maintain global structural consistency [1].

With increased deep learning capabilities, highly sophisticated methods have been proposed. One of the popular techniques is Partial Convolution, which is the method that was tailored for irregular hole inpainting. In this method, the convolution operations are masked and re-normalized in such a way that only valid pixels are considered. As a result, the model can iteratively fill the missing regions, avoiding artifacts caused by invalid inputs [20].

Another major breakthrough is the Contextual Attention mechanism that allows the model to directly access and use information from faraway spatial parts in the image. This technique enhances the capability to restore semantically significant structures by finding and focusing on the appropriate background areas, hence producing more harmonious and beautiful results [21].

More recently, Fourier-based methods have been developed and have achieved the best results. For example, LaMa (Large Mask Inpainting) uses fast Fourier convolutions to not only efficiently handle large holes but also understand the overall context of the image. The method has the ability to apply well to different situations and can produce very good inpainting results even in complicated scenes [22]. Regardless of this progress, issues still persist in perceiving global semantic consistency and dealing with extremely complicated or structured missing regions. Today, further improving the lifelikeness, the speed, the reliability, and the ability to work with various datasets and in real-world situations are the main issues that are being tackled.

4. IMAGE SEGMENTATION AND OBJECT DETECTION METHODS

Image segmentation and object detection are two fundamental problems in computer vision. They are both ways to understand and interpret visual data at a higher level. Image segmentation means separating an image into different parts or pixels with similar characteristics. Object detection means finding and marking the position of certain objects in the image by drawing rectangles and assigning class labels to the objects. Mainly, such tasks are the core of applications like self-driving cars, medical imaging, security, and robots [1].

4.1 Instance and Panoptic Segmentation

Instance segmentation and panoptic segmentation are highly advanced computer vision tasks which go beyond semantic segmentation by giving us a more detailed understanding of scenes. In semantic segmentation, each pixel is labeled with a particular class, whereas instance segmentation even identifies different members of the same class as separate objects, and panoptic segmentation merges the two aspects of semantic and instance-level understanding into a single framework [1].

Instance segmentation finds extensive use in areas that demand individual object identification, such as self-driving cars, robots, and medical imaging. Initially, the methods depended on region proposals; however, nowadays, techniques are mainly built on deep learning architectures. Out of all the methods, Mask R-CNN is considered to be one of the leading models in this branch of research. It is a modification of Faster R-CNN that has an added parallel branch for mask prediction of objects along with bounding boxes and class labels. In this way, the segmentation of objects at a pixel level is achieved with a high degree of accuracy [23].

Panoptic segmentation was initially introduced as a method for combining semantic and instance segmentation into a single problem. A semantic label and instance ID are assigned to each pixel in the image, allowing a whole understanding of the scene. Architectures like Panoptic FPN merge both tasks in a common network, thus making consistency and efficiency improvements in analyzing complex scenes possible [24]. Lately, transformer-based methods have made breakthroughs by modeling long-range interactions and global context that, among other things, allow a better disambiguation of overlapping object and background areas [25]. These are some technological improvements, but the problems of occlusion, computational cost, and real-time performance have not been fully solved. Most of the existing works are aimed at designing more efficient and integrated architectures for deployment in the actual environments [1], [25].

4.2 Transformer Shift Segmentation

Until now, transformer segmentation models have [[only]] scarcely changed the computer vision path, e. g., by substituting convolutional n creating self-attention operations. Compared to classical CNN architectures, these models get long-range dependencies in a much more efficient way, so they are indeed very good for complicated segmentation tasks [1].

Vision Transformers (ViT) represented the first use of a pure transformer architecture for image recognition. However, the training of ViTs demands huge datasets and computational resources as a major disadvantage. Several efficient versions, e. g., Swin Transformer and shifted window-based self-attention with dismantling, were presented by the authors after observing this flip side of ViT in an attempt to lower the computational cost and, at the same time, maintain a very good representational ability [23].

Transformer-based models such as SegFormer and the Swin Transformer-based segmentation millionaires have revealed their potential magnificently in segmentation tasks through the combination of hierarchical feature extraction and global context modeling. The shifted window procedure enables the model to effectively deal with the local and global information, which leads to the improvement in the accuracy of the edges and the consistency of the objects [24]. Transformer shift segmentation models implement methods not only to enhance the capability to deal with difficult scenes but also especially when CNN-based approaches are facing problems in managing long-range dependencies and contextual ambiguities. This kind of model is being increasingly utilized in such areas as medical imaging, autonomous systems, and remote sensing [23], [24].

Transformer-based segmentation models perform very well but still have some challenges, such as being computationally heavy and needing a lot of memory. Researchers are still finding ways to make these models lightweight and combine CNN and transformer models to get better efficiency and scalability [1].

5. METHODOLOGY

In this research, we have resorted to a systematic literature review (SLR) to present a thorough and orderly assessment of the latest innovations in image processing with the help of artificial intelligence (AI) techniques.

Systematic literature review is a recognized scientific inquiry approach that accounts for clear visibility, exact replicability, and minimal distortion in the process of picking and examining scientific articles [25].

The paper gathers worthy papers from well-known scholarly databases such as IEEE Xplore, Springer ScienceDirect, MDPI, and arXiv. These databases are chosen for their wide range of peer-reviewed publications in the areas of artificial intelligence and image processing, which warrant the inclusion of top-notch and trustworthy scientific works [26]. The range of the review is confined to papers that were published from 2015 to 2025. This period is selected in order to reflect the latest progress in artificial intelligence, especially the swift evolution of deep learning techniques and their applications to image processing operations [2]. Specific inclusion criteria are applied to guarantee that the selected studies are relevant and of good quality. The review only considers peer-reviewed journal articles and conference papers. Besides, the chosen studies should primarily deal with the use of artificial intelligence methods in image processing. Moreover, in order to be considered for the review, studies must present experimental evaluations with quantitative results that facilitate an objective comparison of different methods.

In this review paper, we have focused on the image enhancement techniques using AI, including noise removing, super-resolution, and deploring). Different content generation techniques of AI, like CNN models, GAN and transformer models, have evolved over time. The paper investigates the journey step by step in a structured manner. Firstly, the papers chosen are categorized based on the AI methods they have used, such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), or Transfer Learning techniques. Such a sorting criterion makes the discovering of frequently used methods and identifying a general research direction in the area quite straightforward. Furthermore, the cross-comparison here is carried out according to the classic image processing metrics. For instance, Accuracy indicates the correctness of the classification result, the Peak Signal-to-Noise Ratio (PSNR) quantifies the fidelity of the reconstructed image with respect to the original, and the Structural Similarity Index Measure (SSIM) assesses the perceptual similarity between images [27]. These measures are extremely appropriate for objectively differentiating various methods. Ultimately, the introduced plan acts as a nicely structured roadmap to monitor the advancement of image processing technologies pioneered by AI.

6. DATASET USED

Certainly, the quality as well as the variety of datasets have an immense influence on the performance of deep learning models, particularly when dealing with image processing and computer vision. To evaluate the models, benchmark datasets are often utilized, which cover a variety of tasks such as classification, segmentation, super-resolution, and denoising. Among the various datasets available, one of the largest and most prominent datasets in the area of computer vision is ImageNet, which is mainly concerned with image classification and feature learning. It consists of millions of labeled images in thousands of categories that support large-scale deep learning model training [28].

COCO (Common Objects in Context) is a well-known dataset primarily used for image tasks like object detection, segmentation, and captioning. It has a variety of complex real-life scenes, each containing several different objects, and it comes with rich annotations such as bounding boxes and segmentation masks [29]. Places2 is a dataset for scene recognition that mainly focuses on understanding the environment and context portrayed in pictures. The dataset contains millions of images, grouped into a large number of scene categories, covering both indoor and outdoor environments [30].

DIV2K (DIVerse 2K resolution dataset) is primarily used for single-image super-resolution (SISR) tasks. Providing a vast number of 2K-resolution, high-quality images, it is the benchmark dataset for training super-resolution networks such as SRCNN and ESRGAN [31].

BSD68 (Berkeley Segmentation Dataset) is widely recognized as a standard benchmark for image denoising tasks. The dataset contains natural images that serve as test images for different noise removal methods implementation and evaluation at various noise levels [10].

6.1 Evaluation Metrics

For evaluating the effectiveness of image processing models, depending on the particular task, various quantitative metrics are utilized. These are exemplified as follows: Peak Signal-to-Noise Ratio (PSNR). PSNR is one of the most frequent metrics for quantitatively assessing the quality of the restoration of images. The formula for it is:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (1)$$

Where MAX_I is the highest allowable pixel value, while MSE stands for the mean squared error between the original and reconstructed images.

Structural Similarity Index (SSIM). SSIM is used to compare the perceptual similarity of two images, namely, it assesses luminance, contrast, and structural features:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

Structural Similarity Index (SSIM)

Where:

- $\mu_x$and $\mu_y$represent the mean intensities of images $x$and y(luminance)
- $\sigma_x^2$and σ_y^2denote the variances of $x$and y(contrast)
- $\sigma_{xy}$is the covariance between $x$and y(structure)
- $C_1$and $C_2$are small constants used to stabilize the division

Corrected sentence (important fix):

Where $\mu$is the mean, $\sigma^2$is the variance, and $\sigma_{xy}$is the covariance.

Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS measures perceptual similarity using deep neural network features:

$$LPIPS(x,y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \| w_l \odot (f_l(x) - f_l(y)) \|_2^2 \quad (3)$$

Where $f_l$represents deep features extracted from the layer l, and $w_l$are learned weights

Fréchet Inception Distance (FID)

FID evaluates similarity between generated and real image distributions:

$$FID = \| \mu_r - \mu_g \|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \quad (4)$$

Where $mu_r, Sigma_r$are statistics of real images and $mu_g, Sigma_g$are for generated images.

Inception Score (IS)

The Inception Score (IS) is used to evaluate both the quality and diversity of generated images:

$$IS = \exp(\mathbb{E}_x[D_{KL}(p(y \mid x) \| p(y))]) \quad (5)$$

Where $p(y \mid x)$represents the conditional label distribution predicted by a pretrained classifier, and $p(y)$is the marginal distribution over all generated images.

Task-Specific Metrics

Mean Intersection over Union (mIoU) – for image segmentation

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \quad (6)$$

Where:

- TP_i: true positives
- FP_i: false positives

- FN_i: false negatives
- N: number of classes

Average Precision (AP) – for object detection

$$AP = \int_0^1 P(R)\, dR \tag{7}$$

Where $P(R)$denotes precision as a function of recall

7. ANALYSIS AND FUTURE DIRECTIONS

A survey of recent articles on image processing with artificial intelligence shows a departure of the field from traditional image processing methods to the use of deep learning-based approaches. These cutting-edge methods have resulted in significant progress in different image-related operations, including classification, detection, segmentation, enhancement, and even reconstruction.

Convolutional Neural Networks (CNNs) are the basis of the great majority of image processing systems since they have a very strong ability to learn spatial and hierarchical feature representations directly from raw image data. CNNs have been an extremely precise tool in various applications, notably in the fields of image classification and object detection. However, two factors mainly limit the extent of their performance: high computational costs and the need for large labeled datasets. Generative Adversarial Networks (GANs) introduced a powerful generative approach that allowed us to generate synthetic images which are almost indistinguishable from real ones. GANs, in fact, have turned into one of the principal ways to generate images, perform image restoration, and expand datasets. On the other hand, training GANs is a difficult challenge, and they frequently suffer from issues like instability and mode collapse during their learning process.

Transfer learning is often a great solution to the problem of image recognition with small datasets. Training time and computational expense will be drastically reduced when using pre-trained models, and yet the model performance can be maintained at a very high level. However, the effectiveness of the approach largely hinges on the similarity between the source and the target domains. In the last few years, transformer-based models have been able to compete quite effectively with CNNs for image analysis tasks. Their ability to analyze and interpret long-distance relationships and overall context has enabled them to achieve major improvements in performance on the hardest vision tasks. Nevertheless, these models require big datasets and computational power, which can limit their deployment in practice in some scenarios. Hybrid models combining two or more deep learning architectures, namely CNNs, Transformers, and GANs, have proven to be very powerful in generating very precise and reliable results. Hybrid models aim at combining local feature extraction, global context understanding, and image enhancement in a single framework. In summary, AI-powered image processing has become a very potent and versatile tool. However, some of the biggest challenges, such as huge computational requirements, dependence on data, and lack of interpretability of models, remain as barriers to the full deployment of AI in real-world scenarios. A brief comparative of various AI technologies is presented in the table below.

AI-assisted image processing is a field that is rapidly developing. One of the possible directions of this technology is the creation of hybrid architectures that integrate CNNs, Transformers, and GANs to realize larger gains in both efficiency and accuracy. Broadly speaking, the architecture that has been developed for that purpose is outlined as follows: Input Image Preprocessing, CNN Feature Extraction, Transformer Attention Module, GAN Enhancement, and Output. This architecture is capable of learning features at a deeper level and upgrading the image quality and application to real-world scenarios with a very high level of robustness. Such a system can explore the features of the images more intensively, upgrading the image quality and robustness for real-life use.

Table 1. Comparative Discussion of AI Techniques in Image Processing

Technique	Main Role	Advantages	Limitations	Typical Applications
CNNs	Feature extraction and classification	High accuracy, strong feature learning	High computation, needs large datasets	Image classification, object detection
GANs	Image generation and beyond	photo-realistic images	Training instability, difficult optimization	Image synthesis, restoration
Transfer Learning	Model reuse for new tasks	Reduces training time, works with small datasets	Limited adaptability to new domains	Medical imaging, general classification
Transformers	Global context learning	High accuracy, captures long-range dependencies	Requires large data and computation	Advanced image recognition tasks

Technique	Main Role	Advantages	Limitations	Typical Applications
Hybrid Models	Combine multiple methods	Improved performance, flexible	Complex design and tuning	Complex image processing systems

8. CONCLUSION

Artificial intelligence has drastically altered the landscape of image processing. Deep learning methods, for instance, have continued to dominate research, but at the same time, they have become the real solutions in the industry. CNNs are known for their ability to automatically identify features and recognize objects with such high accuracy, whereas GANs, on the other hand, deal with image generation and restoration. Transformers, which depend on the self-attention mechanism, are capable of getting the global context which in turn results in more, are capable of getting the global context which in turn results in more capable image analysis systems. Yet, regardless of these great changes, a number of problems are still there, such as features that require huge computation, heavy dependence on large datasets, and very limited interpretability, which together lead to making the practical use of these systems nearly impossible. The main aim of research in the future should be devoted to the creation of AI models, e. g., efficient lightweight ones, explainable and also hybrid architectures that can do best by combining CNNs, GANs, and Transformers for scalability and better performance.

9. REFERENCES

[1] Szeliski R. Computer Vision: Algorithms and Applications. Springer, New York, USA, 2010. DOI: 10.1007/978-1- 84882-935-0.

[2] LeCun Y., Bengio Y., Hinton G. "Deep learning," *Nature*, 2015, 521(7553), 436–444. DOI: 10.1038/nature14539.

[3] Dosovitskiy A., Beyer L., Kolesnikov A., et al. "An image is worth 16×16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021. DOI: 10.48550/arXiv.2010.11929.

[4] Litjens G., Kooi T., Bejnordi B.E., et al. "A survey on deep learning in medical image analysis," *Medical Image Analysis*, 2017, 42, 60–88. DOI: 10.1016/j.media.2017.07.005.

[5] Krizhevsky A., Sutskever I., Hinton G.E. "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012, 25, 1097–1105.

[6] Goodfellow I., Pouget-Abadie J., Mirza M., et al. "Generative adversarial nets," *Advances in Neural Information Processing Systems*, 2014, 27, 2672–2680.

[7] He K., Zhang X., Ren S., Sun J. "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 770–778. DOI: 10.1109/CVPR.2016.90.

[8] Long J., Shelhamer E., Darrell T. "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[9] Pan S.J., Yang Q. "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10), 1345–1359. DOI: 10.1109/TKDE.2009.191.

[10] Zhang Q., Zhu S.C. "Visual interpretability for deep learning: A survey," *IEEE Transactions on Big Data*, 2021, 7(1), 76–95. DOI: 10.1109/TBDATA.2018.2876803.

[11] Gonzalez R.C., Woods R.E. Digital Image Processing, 4th ed. Pearson, Hoboken, NJ, USA, 2018.

[12] Zhang K., Zuo W., Chen Y., Meng D., Zhang L. "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, 2017, 26(7), 3142–3155. DOI: 10.1109/TIP.2017.2662206.

[13] Ledig C., Theis L., Huszár F., et al. "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 4681–4690. DOI: 10.1109/CVPR.2017.19.

[14] Vaswani A., Shazeer N., Parmar N., et al. "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017, 30, 5998–6008.

[15] Dong C., Loy C.C., He K., Tang X. "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2), 295–307. DOI: 10.1109/TPAMI.2015.2439281.

[16] Wang X., Yu K., Wu S., et al. "ESRGAN: Enhanced super-resolution generative adversarial networks," *ECCV Workshops*, 2018. DOI: 10.1007/978-3-030-11021-5_5.

[17] Liang J., Cao J., Sun G., et al. "SwinIR: Image restoration using Swin Transformer," *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2021, 1833–1844. DOI: 10.1109/ICCVW54120.2021.00210.

[18] Ronneberger O., Fischer P., Brox T. "U-Net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, 2015, 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[19] Kupyn O., Budzan V., Mykhailych M., et al. "DeblurGAN: Blind motion deblurring using conditional adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 8183–8192. DOI: 10.1109/CVPR.2018.00854.

[20] Liu G., Reda F.A., Shih K.J., et al. "Image inpainting for irregular holes using partial convolutions," *European Conference on Computer Vision*, 2018, 85–100. DOI: 10.1007/978-3-030-01252-6_6.

[21] Yu J., Lin Z., Yang J., et al. "Generative image inpainting with contextual attention," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 5505–5514. DOI: 10.1109/CVPR.2018.00577.

[22] Suvorov R., Logacheva E., Plosskaya A., et al. "Resolution-robust large mask inpainting with Fourier convolutions (LaMa)," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022, 4292–4302. DOI: 10.1109/CVPRW56347.2022.00478.

[23] Kirillov A., He K., Girshick R., Rother C., Dollár P. "Panoptic segmentation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 9404–9413. DOI: 10.1109/CVPR.2019.00963.

[24] Liu Z., Lin Y., Cao Y., et al. "Swin Transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 10012–10022. DOI: 10.1109/ICCV48922.2021.00986.

[25] Xie E., Wang W., Yu Z., et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, 2021, 34, 12077–12090. DOI: 10.48550/arXiv.2105.15203.

[26] Kitchenham B. "Procedures for performing systematic reviews," Keele University Technical Report TR/SE-0401, 2004.

[27] Brereton P., Kitchenham B.A., Budgen D., Turner M., Khalil M. "Lessons from applying systematic literature reviews in software engineering," *Empirical Software Engineering*, 2007, 12(6), 571–583. DOI: 10.1007/s10664-007-9024-1.

[28] Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. "ImageNet: A large-scale hierarchical image database," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 248–255. DOI: 10.1109/CVPR.2009.5206848.

[29] Lin T.Y., Maire M., Belongie S., et al. "Microsoft COCO: Common objects in context," *European Conference on Computer Vision*, 2014, 740–755. DOI: 10.1007/978-3-319-10602-1_48.

[30] Zhou B., Lapedriza A., Xiao J., Torralba A., Oliva A. "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(6), 1452–1464. DOI: 10.1109/TPAMI.2017.2723009.

[31] Agustsson E., Timofte R. "NTIRE 2017 challenge on single image super-resolution: Dataset and study," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, 126–135. DOI: 10.1109/CVPRW.2017.15.

Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.54
https://jatuc.uoturath.edu.iq/

تقييم التطورات في معالجة الصور من خلال أساليب الذكاء الاصطناعي: مراجعة

امل عباس كاظم[1]*، وداد عبد الخضر ناصر[2]، سفانة حيدر عباس[3]

[3,2,1] قسم علوم الحاسبات، الجامعة المستنصرية كلية التربية، بغداد، العراق.

* *البريد الإلكتروني للتواصل:* abbas.amal@uomustansiriyah.edu.iq

الملخص

أُدرك أن الذكاء الاصطناعي يمثل قوة رئيسة في إحداث ثورة في مجال معالجة الصور. وبناءً على ذلك، تستعرض هذه الورقة أحدث تقنيات معالجة الصور، ولا سيما تقنيات الذكاء الاصطناعي التي تشمل أساليب التعلم الآلي والتعلم العميق. وقد أسهمت التقنيات الحديثة القائمة على الذكاء الاصطناعي، مثل الشبكات العصبية الالتفافية (CNNs) والنماذج القائمة على المحولات(Transformers) ، في تحسين مستوى أداء عدد من مهام معالجة الصور بدرجة كبيرة، ومنها تصنيف الصور، وكشف الأجسام، وتحسين الصور، والتعرّف على الأنماط. كما تعد تقنيات التصوير الطبي المتقدمة، والأمن، والأنظمة ذاتية التشغيل، وتقنيات الوسائط المتعددة من أبرز مجالات تطبيق معالجة الصور المدعومة بالذكاء الاصطناعي التي تناولتها هذه المراجعة.

وتسلط الورقة الضوء على أبرز التحديات، مثل الاعتماد على مجموعات بيانات ضخمة، وارتفاع التكلفة الحاسوبية، وصعوبة تفسير النماذج وفهم آليات عملها. وأخيرًا، تناقش الاتجاهات الواعدة للبحوث المستقبلية التي تهدف إلى تطوير نماذج ذكاء اصطناعي كفؤة وقابلة للتوسع والتفسير. وبوجه عام، تقدم هذه المقالة مراجعة شاملة لدور الذكاء الاصطناعي بوصفه أحد أكثر عوامل التغيير تأثيرًا في مجال معالجة الصور.

الكلمات المفتاحية

الذكاء الاصطناعي، معالجة الصور، التعلم الآلي، التعلم العميق، الشبكات العصبية التلافيفية، نماذج المحولات.