IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, June 2023

A Survey on the Classification of Research Paper

Disha Deotale[1], Anisha Kharare[2], Sagar Nagarale[3], Vasundhara Gajbhiye[4]

Department of Information Technology[1,2,3,4]

G. H. Raisoni Institute of Engineering and Technology, Pune, India

Abstract: *Image recognition has become a fundamental research area in computer vision and artificial intelligence. This survey paper provides a comprehensive overview of state-of-the-art image recognition techniques, highlighting their advancements, challenges, and applications. It explores the challenges faced in image recognition, such as domain adaptation, limited labeled data, interpretability, and robustness to variations. It also discusses practical applications of image recognition in various domains, such as healthcare, autonomous driving, surveillance, and augmented reality.*

Keywords: *Image recognition*

I. INTRODUCTION

The history of text can be traced back over thousands of years. Rich and precise semantic information carried by text is important in a wide range of vision-based application scenarios. Therefore, text recognition in natural scenes has been an active research field in computer vision and pattern recognition. Recognizing text in natural scenes has attracted great interest from academia and industry in recent years because of its importance and challenges. Text in natural scenes can provide rich and precise information, which is beneficial for understanding the scene. Automatic recognition of text in natural scenes is economically viable in the era of big data, which attracts researchers and practitioners. Extraction and recognition of text from images is an important step in building efficient indexing and retrieval systems for multimedia databases.

- **Background:** Unlike OCR in scanned documents, text in natural scenes can appear on anything (e.g., signboards, walls, or product packaging). Therefore, scene text images may contain very complex backgrounds. Moreover, the texture of the background can be visually similar to the text, which causes additional challenges for recognition.
- **Forms:** Text in scanned documents is usually printed in a single color with regular font, consistent size, and uniform arrangement. In natural scenes, text appears in multiple colors with irregular fonts, different sizes, and diverse orientations. The diversity of text makes STR more difficult and challenging than OCR in scanned documents.
- **Noise:** Text in natural scenes is usually distorted by noise interference, such as nonuniform illumination, low resolution, and motion blurring. Imperfect imaging conditions cause failures in STR.
- **Access:** The scanned text is usually frontal and occupies the main part of the image. However, scene text is captured randomly, which results in irregular deformations (such as perspective distortion). Various shapes of text increase the difficulty of recognizing characters and predicting text strings.

II. BACKGROUND

Background extraction is the process of separating foreground objects from the background of an image. Text extraction from images, also known as optical character recognition (OCR), is the process of converting text contained within an image into machine-readable text. Special issues with text extraction include text orientation, text size, and scaling, font and style variations, background interference, handwritten or stylized text, and multilingual text. To address these issues, advanced algorithms and techniques such as deep learning and convolutional neural networks have been used to improve the accuracy and robustness of OCR systems.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, June 2023

2.1 TEXT IN IMAGES

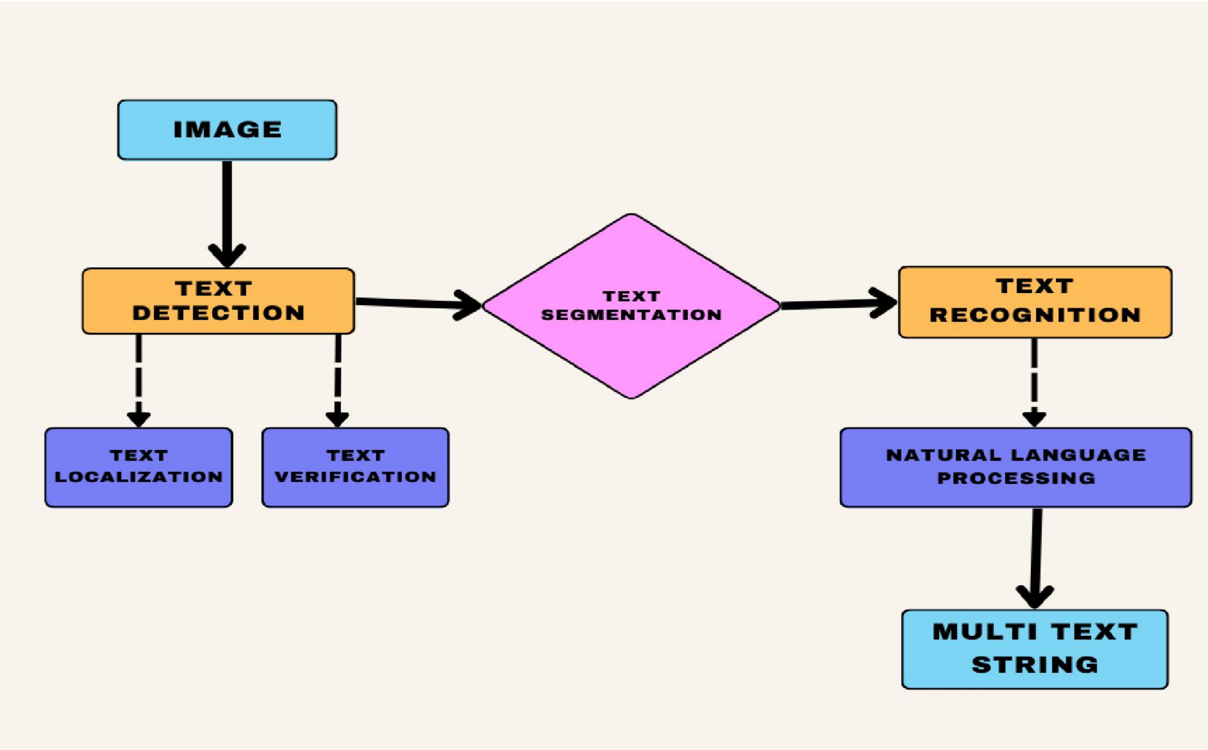
Optical Character Recognition (OCR) is a technology used to convert text from images or scanned documents into machine-readable text. It involves pre-processing the image, image segmentation, text extraction, text cleaning, and post-processing. Pre-processing involves tasks such as resizing, denoising, and adjusting the contrast or brightness. Image segmentation involves thresholding, edge detection, or contour extraction. Text extraction involves applying OCR algorithms to the image pixels and converting them into characters or words.

Text cleaning involves spell-checking, punctuation correction, and noise removal. Post-processing involves language identification, text formatting, or further text analysis.

2.2 FUNDAMENTAL PROBLEMS AND SPECIAL ISSUES WITH TEXT

Text can be ambiguous and its meaning can depend on the context in which it is used. Polysemy, homonyms and homophones, synonyms, and ambiguous pronouns can all lead to confusion or miscommunication. Understanding these issues is essential for accurate text comprehension and information retrieval. Resolving pronouns correctly is essential for understanding the intended meaning. Idiomatic expressions, sarcasm, irony, named entity recognition, multilingual text, and text data quality are all challenges that require advanced natural language processing techniques. Researchers and practitioners are constantly developing algorithms and models to improve text processing systems.

Fig. 01 Illustration of Problems



- **Text Detection:** The function of text detection, is to determine whether text is present using localization and verification procedures. As a basis of an end-to-end system, it provides precise and compact text instance images for text recognition. Text detection approaches can be roughly categorized as regression-based and instance segmentation-based methods.
- **Text Localization:** Text localization aims to localize text components precisely and to group them into candidate text regions with as little background as possible. Early text localization methods are based on low-level features, such as color, gradient, stroke width transform, maximally stable extremal regions (MSER), canny detectors, and connected component analysis. Most of the current methods are based on deep neural networks.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

- **Text Verification:** Text verification aims at verifying the text candidate regions as text or non-text. It is usually used after text localization to filter the candidate regions because text localization sometimes introduces false positives. Approaches to text verification include prior knowledge, support vector machine (SVM) classifier, and conditional random fields (CRFs). Recent works used a convolution neural network (CNN) to improve text/non-text discrimination.
- **Text Segmentation:** Text segmentation has been identified as one of the most challenging problems. It includes text line segmentation and character segmentation. The former refers to splitting a region of multiple text lines into multiple sub-regions of single text lines. The latter refers to separating a text instance into multiple regions of single characters. Character segmentation was typically used in early text recognition approaches.
- **Text Recognition:** Text recognition translates a cropped text instance image into a target string sequence. It is an important component of an end-to-end system, which provides credible recognition results. Traditional text recognition methods rely on hand-crafted features, such as histograms of oriented gradient descriptors, connected components, and stroke width transform. Most recent studies have used deep learning encoder-decoder frameworks.
- **End-to-end System:** Given a scene text image, an end-to-end system can directly convert all text regions into the target string sequences. It usually includes text detection, text recognition, and postprocessing. The construction of real-time and efficient end-to-end systems has become a new trend in recent years. Some researchers, interpret text detection and text recognition as two independent subproblems, which are combined to construct an end-to-end system. Another approach is to jointly optimize text detection and text recognition by sharing information.

III. BASIC PROCESS OF A CHARACTER RECOGNITION SYSTEM

The basic process of a character recognition system typically involves the following steps:

- **Pre-Processing:** Pre-processing is the first and the major step of OCR software. OCR (Optical Character Recognition) software is valuable for businesses, organizations, and individuals in various industries. OCR software reads text from physical documents, images, or other sources and converts them into digital formats.At this stage, certain operations are performed on the scanned image i.e. de-skew, converting an image from color to black and white, cleaning up non-glyph boxes and lines, identifying columns, paragraphs, and captions as different blocks, and normalization.
- **Segmentation:** The aim of image segmentation is to provide labels to each pixel in an image such that pixels with the same label share certain visual characteristics. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc) in images. The method of segmentation used in this is edge detection
- **Feature Extraction**: The aim of feature extraction is to capture the essential characteristics of the symbols, and it has been accepted that this is one of the biggest problems of pattern recognition. In this, the approach is to extract certain features that characterize the symbols but leaves out the unimportant attribute. The Selection of the appropriate feature-extracting method is probably one of the most important factors in achieving high recognition performance.
- **Classification and Recognition**: The classifying and identifying of each character and assigning to it the correct character class is called classification. In this stage, the decision-making of a recognition system uses all the features extracted in the earlier stage.
- **Post Processing**: It is the final step of the recognition system being discussed. It prints the corresponding characters which were recognized in the structured text form which is done by the calculation of equivalent ASCII values using the recognition index.

IV. METHODOLOGY FOR IMAGE RECOGNITION TECHNIQUE

Image recognition techniques enable computers to interpret and analyze visual data, enabling applications such as object detection, image classification, and semantic segmentation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

4.1 Traditional Methods

Traditional methods for image recognition typically involve a combination of image processing techniques and machine learning algorithms.

- **Feature-based methods (SIFT, HOG, etc):** Feature-based methods, such as SIFT, HOG, and SURF, are widely used in computer vision for image recognition and object detection tasks. SIFT is a feature detection and description algorithm that identifies key points or interest points in an image and is invariant to scale, rotation, and illumination changes. HOG is a feature descriptor technique that captures the distribution of gradient orientations in an image. SURF is a feature detection and description algorithm that is robust to scale and rotation changes and has faster computation. ORB is a fusion of two techniques - FAST and BRIEF.
- **Bag-of-visual-words models:** The Bag of Visual Words (BoVW) model is a popular approach for image representation and recognition in computer vision. It consists of four steps: feature extraction, feature quantization, feature encoding, and classification or retrieval. The BoVW model captures the distribution of visual words or features in an image, ignoring spatial information. For tasks that require more detailed information, other techniques such as convolutional neural networks (CNNs) are typically used.
- **Template matching techniques:** Template matching is a technique used in computer vision to locate instances of a template image within a larger target image. It involves comparing the template image with different parts of the target image to find the best match. Common template matching techniques include Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC), Zero-Mean Normalized Cross-Correlation (ZNCC), and Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). These techniques are simple and intuitive but can be sensitive to changes in lighting conditions, occlusions, and variations in object appearance. More advanced techniques, such as feature-based methods or deep learning approaches, are often used for more challenging image recognition tasks.

4.2 Deep Learning-Based Methods

Deep learning-based methods have revolutionized image recognition, with CNNs being the primary architecture used. Here are key methods for image recognition.

- **Convolutional Neural Network:** Convolutional Neural Networks (CNNs) are a class of deep learning models designed for processing and analyzing structured grid-like data, such as images. They have revolutionized various computer vision tasks, such as image classification, object detection, and image segmentation. The key components and operations in a CNN are convolutional layers, activation functions, pooling layers, and fully connected layers. CNNs are trained using backpropagation, pre-trained models, and transfer learning. They excel at learning hierarchical representations of visual data, discovering important features and patterns from raw pixel values, and are widely used in image analysis and computer vision. They have achieved state-of-the-art performance in benchmarks and are widely used in practical applications.
- **Transfer learning and pre-trained models:** Transfer learning and pre-trained models are powerful techniques in deep learning that leverage knowledge gained from pre-training on large-scale datasets to solve new tasks with limited data. Pre-trained models are deep learning models trained on large-scale datasets to learn generic features and patterns. Transfer learning involves taking a pre-trained model and adapting it to a new task, while fine-tuning involves further training the pre-trained model on the new task-specific data. Transfer learning and pre-trained models enable the utilization of deep learning models' learned representations for related tasks, such as object recognition, image classification, and image segmentation. Feature extraction and domain adaptation are useful for small and similar datasets, while adversarial training and domain adaptation algorithms are used to align the source and target domains during training.
- **Region-based CNNs (R-CNN, Fast R-CNN, Faster R-CNN):** R-CNN, Fast R-CNN, and Faster R-CNN are a series of deep learning-based object detection models that revolutionized the field of object detection. R-CNN consists of region proposal generation, CNN feature extraction, object classification and localization, and fine-tuning. Fast R-CNN improved upon R-CNN by introducing Region of Interest (RoI) pooling. Faster R-CNN is a widely adopted and highly effective object detection framework due to its improved accuracy and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

efficiency. It features a Region Proposal Network (RPN) that generates object proposals directly from the CNN feature maps, region classification and localization, and unified training. It has achieved state-of-the-art results on various object detection benchmarks.

- **Fully Convolutional Networks (FCNs):** Fully Convolutional Networks (FCNs) are a class of deep learning models designed specifically for semantic segmentation tasks. They utilize transposed convolutions, skip connections, end-to-end training, transfer learning, and dilated (atrous) convolutions to increase the receptive field and capture larger context information. FCNs have revolutionized the field of semantic segmentation by enabling efficient and accurate pixel-wise classification. FCNs can be trained end-to-end using pixel-wise cross-entropy loss, transfer learning, and dilated convolutions. They have achieved state-of-the-art performance on several benchmark datasets and have become a key component in many computer vision applications.
- **Attention mechanisms in image recognition:** Attention mechanisms are used in image recognition to help models focus on relevant regions or features. Spatial attention mechanisms assign importance weights to different spatial locations based on their relevance to the task. Channel attention mechanisms focus on the importance of different channels within a convolutional neural network and learn to assign weights to each channel based on its relevance. Attention mechanisms can be used to enhance image recognition models by emphasizing informative channels and suppressing less relevant ones. Self-Attention mechanisms enable models to capture long-range dependencies and relationships between different parts of an image, while Multi-Head Attention combines multiple attention heads in parallel to capture distinct patterns or relationships. By incorporating attention mechanisms, models can better handle complex and cluttered scenes, improve localization accuracy, and achieve state-of-the-art performance on challenging datasets.
- **Generative Adversarial Networks (GANs) in image synthesis and recognition:** GANs are used for image synthesis and recognition tasks, consisting of two neural networks, a generator and a discriminator, which are trained together in an adversarial setting. GANs have achieved remarkable success in generating realistic and high-quality images and have been utilized for tasks such as image-to-image translation and image super-resolution. GANs are used for image-to-image translation, image recognition, and generative models. CGANs and CycleGANs are popular GAN architectures for image-to-image translation, while GANs can also be used for image synthesis. GANs have advanced the fields of image synthesis and recognition, generating realistic images, performing image-to-image translation, and augmenting training data for image recognition.

V. SUMMARY AND FUTURE SCOPE

Future directions in image recognition are driven by emerging technologies and ongoing research efforts. Robustness and generalization, explainability and interpretability, lifelong learning and Adaptability, small data and few-shot learning, and multi-modal and cross-modal recognition are key areas of focus. These areas will enable a more comprehensive understanding and analysis of complex visual scenes. Continual Annotation and Active Learning, Ethical and Fair Image Recognition, Real-time and Efficient Inference, Large-scale Image Recognition, and Domain-Specific Image Recognition are the future directions and open challenges in image recognition. Advancements in machine learning, computer vision, and data collection techniques will drive progress in the field, leading to more accurate, robust, and interpretable image recognition systems.

- **Hybrid approaches and multimodal image recognition:** Hybrid approaches and multimodal image recognition involve combining information from multiple sources to improve performance and understanding. Hybrid approaches and multimodal recognition techniques aim to improve the accuracy, robustness, and understanding of visual data by combining information from multiple modalities.
- **Lifelong and continual learning in image recognition:** Lifelong and continual learning in image recognition involves addressing catastrophic forgetting, regularization, replay, task ordering, knowledge distillation, and transfer learning to retain knowledge and adapt to new tasks. Lifelong and continual learning in image recognition aims to enable models to learn from a continuous stream of data, adapt to new tasks, and retain knowledge from previous tasks.

- **Addressing the bias and fairness issues in image recognition:** Approaches to address bias and fairness in image recognition include diverse and representative training data, data pre-processing and augmentation, bias detection and mitigation, and fairness-aware training. Addressing bias and fairness in image recognition requires a combination of diverse and representative training data, algorithmic approaches, transparency, inclusive development practices, and continuous monitoring.
- **Privacy and security concerns in image recognition systems:** Implement strong data privacy measures, consent and transparency, data minimization, secure storage and transmission, and robust authentication and access controls to mitigate privacy and security risks in image recognition. Implement strong passwords, multi-factor authentication, and role-based access control to protect against unauthorized access and data breaches, detect and mitigate adversarial attacks, conduct regular security audits, comply with privacy regulations, and use and deploy responsibly.

VII. ACKNOWLEDGEMENT

This research is conducted at G. H. Raisoni Institute of Engineering and Technology, Pune, We would like to thank Prof. Disha Deotale for her support for the research work. We would also like to thank Dr. Rachna Sable, HOD, Department of Computer Science and Information Technology for her consistent guidance, support, and encouragement throughout the research work.

REFERENCES

[1]. Vijay Laxmi Sahu, Babita Kubde,” Offline Handwritten Character Recognition Techniques using Neural Network: A Review “, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064

[2]. R. Seiler M. Schenkel E Eggimann,”Off-Line Cursive Handwriting Recognition Compared with On-Line Recognition”, 1015-4651/96 $5.00 0 1996 IEEE Proceedings of ICPR '96.

[3]. XIAOXUE CHEN, LIANWEN JIN, YUANZHI ZHU, CANJIE LUO, and TIANWEI WANG. Text Recognition in the Wild: A Survey. J.ACM, 1. (1),2020.

[4]. R.-C Chen, “Automatic license plate recognition via sliding-window darknet-YOLO deep learning,” *Image Vis. Comput.,* vol. 87, pp. 47-56, Jul. 2019.

[5]. X. Liu, G. Meng, and C. Pan, “Scene text detection and recognition with an advance in deep learning: A survey,” *Int. J. Document Anal.* Recognit., vol. 22, no.2, pp. 143-162, jun. 2019.

[6]. Achakulvisut, T., Acuna, D. E., Ruangrong, T., & Kording, K. (2016). Science concierge: A fast content-based recommendation system for scientific publication. *PLoS ONE, 11(7), e0158423.*

[7]. Rokus Arnold, Poth Miklos, ” Character Recognition Using Neural Networks”, 11th IEEE International Symposium on Computational Intelligence and Informatics • 18–20 November, 2010.

[8]. Tasweer Ahmad, Ahlam Jameel, Dr. Balal Ahmad, ”Pattern Recognition using Statistical and Neural Techniques”, 978-1-61284-941-6/11/$26.00 20 11 IEEE.

[9]. Jon Almazán, Albert Gordo, Alicia Fornés, and Ernest Valveny. 2014. Word spotting and recognition with embedded attributes. IEEE Trans. Pattern Anal. Mach. Intell 36, 12 (2014), 2552–2566

[10]. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In Proceedings of ICLR.

[11]. Xiang Bai, Mingkun Yang, Pengyuan Lyu, Yongchao Xu, and Jiebo Luo. 2018. Integrating scene text and visual appearance for fine-grained image classification. IEEE Access 6 (2018), 66322–66335.

[12]. Yong Cheng. 2019. Semi-supervised learning for neural machine translation. In Joint Training for Neural Machine Translation. 25–40

[13]. Zhanzhan Cheng, Fan Bai, Yunlu Xu, Gang Zheng, Shiliang Pu, and Shuigeng Zhou. 2017. Focusing attention: Towards accurate text recognition in natural images. In Proceedings of ICCV. 5086–5094.

[14]. Li Chengxiong, Ding Yuehua, Wen Guihua. Application of SVM – KNN Combination Improvement Algorithm on Patent Text Classification[J]. Computer Engineering and Applications, 2006, 42(20),193-195.

[15]. Li Shengzhen, Wang Jianxin, Qi Jiandong. Zutomated Categorization of Patent Based on Backpropagation Network[J]. Computer Engineering and Design,2010,31(23),5075-5078.

[16]. Lai K K, Wu S J., Using the Patent Co–citation Approach to Establish a New Patent Classification System[J]. Information Processing and Management, 2005, 41(2),313-330.

[17]. D. Deotale M. Verma, P. Suresh, Divya Srivastava, M. Kumar, S. Kumar Jangir. 2022. Analysis of Human Activity Recognition Algorithm Using Trimmed Video Datasets