IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 November 2023, accepted 12 January 2024, date of publication 15 January 2024, date of current version 29 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354706

RESEARCH ARTICLE

Synergistic Integration of Transfer Learning and Deep Learning for Enhanced Object Detection in Digital Images

SAFA RIYADH WAHEED[1,2], NORHAIDA MOHD SUAIB[1], MOHD SHAFRY MOHD RAHIM[3], AMJAD REHMAN KHAN[4], (Senior Member, IEEE), SAEED ALI BAHAJ[5], AND TANZILA SABA[4], (Senior Member, IEEE)

[1]Faculty of Engineering, School of Computing, Universiti Teknologi Malaysia, Skudai, Johor Bahru 81310, Malaysia
[2]Computer Techniques Engineering Department, College of Technical Engineering, Islamic University, Najaf 54001, Iraq
[3]Media and Games Innovation Centre of Excellence, UTM-IRDA Digital Media Centre, Institute of Human Centred Engineering, Universiti Teknologi Malaysia, Skudai, Johor 81310, Malaysia
[4]Artificial Intelligence and Data Analytics Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia
[5]MIS Department, College of Business Administration, Prince Sattam bin Abdulaziz University, AlKharj 11942, Saudi Arabia

Corresponding author: Saeed Ali Bahaj (bahajsaeedali@gmail.com)

ABSTRACT Presently, the world is progressing towards the notion of smart and secure cities. The automatic recognition of human activity is among the essential landmarks of smart city surveillance projects. Moreover, classifying group activity and behavior detection is complex and indistinct. Consequently, behavior classification systems reliant on visual data hold expansive utility across a spectrum of domains, including but not limited to video surveillance, human-computer interaction, and the safety infrastructure of smart cities. However, automatic behavior classification poses a significant challenge in the context of live videos captured by the smart city surveillance system. In this regard, the use of pictures with pre-trained convolution neural networks (CNNs)-assisted transfer learning (TL) has emerged as a potential technique for deep neural networks (DNNs) object detection., resulting in increased performance in localization for smart city surveillance. Against this backdrop, this paper explores various strategies to develop advanced synthetic datasets that could enhance accuracy when trained with modern DNNs for object detection (mAP). TL was employed to address the limitation of DL that necessitates a huge dataset. The KITTI datasets were used to train a contemporary DNN single-shot multiple box detector (SSMD) in TensorFlow. A variety of metrics were employed to assess the efficacy of the novel automated Transfer Learning (TL) system within a real-world context, specifically designed for object detection within the DL framework (referred to as OD-SSMD). The results unveiled that this developed system outperformed preceding investigations, demonstrating superior performance. Notably, it exhibited the remarkable capability to autonomously discern and pinpoint various attributes and entities within digital images, effectively identifying and localizing each item present within the images.

INDEX TERMS Object detection, TL, DL, SSMD, CNN, VGG16, smart city, security, technological development.

I. INTRODUCTION

The advancements in automation have greatly expanded the range of view of modern computer vision (CV) systems, enabling their use in various specialized industries such as robotics, manufacturing, building automation, intelligent sensors, medical imaging, food processing, and autonomous driving. This progress has been facilitated by the development of multi-core architectures and improved processor designs, which have enabled faster clock rates and GPU-assisted parallel data processing, which has significantly increased operational bandwidth. Additionally, Deep convolutional

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Tucci.

neural network (DCNN) architectures can now perform object identification and classification tasks.

Neural networks (NNs) have garnered extensive acclaim within computer technology due to their remarkable ability to generalize extensive datasets with minimal processing complexity. NNs can approximate any designated nonlinear function with exceptional precision, rendering them exceedingly self-sufficient and adept at achieving robust generalization. Nevertheless, the task of image categorization using Deep Convolutional Neural Networks (DCNNs) continues to pose significant challenges, primarily because the efficacy of CNNs hinges upon layer-specific attributes. In the context of object recognition, CNNs employ convolution to harness multiple feature maps within each layer, thereby unveiling a novel, spatially invariant feature ensemble. These networks also use tailored learning mechanisms for distinct purposes, including object localization.

To surmount the issues surrounding existing computer vision systems, comprehensively produced images can be applied to increase the overall quality of the training data. Transfer learning (TL) is a technique that allows information to be transferred from one subject to another, and synthetic datasets-based pre-trained models can guarantee successful transfer of features. However, using fake datasets can be problematic. Open-source models were evaluated for their efficiency in producing images with low-level signals such as textures, poses, and context were employed to train DCNNs, and the results demonstrated that low cue constancy and photorealism levels did not interfere significantly with training networks that used synthetic data.

Multiple tiers of Convolutional Neural Networks (CNN) can be strategically employed to achieve feature transferability. This transferability signifies the inherent versatility of the initial layers, which can be effectively leveraged for various object recognition tasks. This concept underscores the potential for fine-tuning features and crafting synthetic datasets with specific attributes to facilitate optimal training. It emphasizes the importance of a cooperative approach to dataset augmentation, involving the integration of synthetic objects into the real-world environment to enrich dataset diversity. These methodologies, as outlined above, share a common thread in harnessing synthetic datasets across supplementary resource categories to enhance the accuracy of Deep Neural Networks (DNNs).

This paper conducts a comprehensive assessment of the performance of various single-shot multiple box detector (SSMD) models, all of which have been meticulously trained. These models are rigorously tested using diverse datasets that have been meticulously prepared in a consecutive manner. Notably, the evaluation focuses on practical images that have never been encountered in prior detection tasks, spanning a spectrum of object types. The study delves into the effectiveness of various data synthesis techniques applied to these datasets, comparing their outcomes with those of other datasets used for accurate feature classification. At its core, the primary goal of this research is to achieve a seamless transfer of learned representations or features from synthetic to real-world domains. Such an achievement significantly augments the proposed system's overall performance. To culminate, the study also thoroughly examines both the quantity and diversity of synthetic datasets, coupled with the fine-tuning of deep neural network (DNN) hyperparameters. This comprehensive approach optimizes the entire object identification pipeline for effective Transfer Learning (TL).

II. RELATED WORKS

Through a comprehensive grasp of the contextual intricacies surrounding the issue at hand, it can be inferred that object detection studies still necessitate further exploration [15], [16]. The greatest obstacle in object detection is to represent an object with efficient and effective feature extraction techniques for object representation [17], [18], which aims to enhance the precision of detection performance. Object detection finds wide-ranging application across diverse real-world scenarios, encompassing domains like autonomous vehicular navigation, robotic vision systems, and advanced video surveillance setups [19], [20], [21]. Before the emergence of CNNs, the deform parts model (DPM) [22] and chosen to search [23] showed similar performance. However, the advent of increased recurrence CNN (R-CNN) [24] merged the chosen search regions proposal and post-classification CNN, resulting in the emergence of regions' proposals for classification-based object-detection systems. Subsequently, several enhancements were made to the original R-CNN approach, including the use of image collection categorization. In order to elevate both the quality caliber and speed of post-classification processes, as well as the introduction of the SPPnet, which the original R-CNN technique was greatly expedited. [25]. This was made possible by introducing pyramid pool layers, which enabled the classification of the layer by reusing the spatial features obtained from the features map created with different images resolution.

The Rapid R-CNN [26] expanded the SPPnet by modifying all layers via the loss minimization continuously bound boxes regressions and confidences proposed in the multiple boxes [27] to learn objects. Subsequently, deep neural networks (DNNs) improved proposal generation features [28], [29]. However, most recent works have abandoned the use of low-level image features-based proposals like multibox [27], [30] in favor of generating proposals directly from a separate DNN [31], [32], [33]. This produces a more multifaceted system requiring two correlated NNs training, significantly enhancing detection accuracy [34]. The current strategy falls under this category, as it lacks the proposal step contains only default boxes. While default boxes offer greater flexibility compared to current systems, it's noteworthy that they encompass a spectrum of aspect ratios and are amenable to deployment across diverse scales at each feature location [35].

To train a model to recognize an item, computer vision models are designed to learn which pixel patterns relate to the object of interest [36]. As a result, this study encounters various obstacles, including improving retrieval accuracy and refining object descriptors and feature extraction stages [37]. The three components of a complete object detection system are feature extraction, object recognition, and object localization. Object detection has essential applications in various domains, as previously mentioned. Consequently, this study considers the aforementioned extra qualities in addition to the comparable criticality, and the term ''usability'' refers to the system's ability to recognize an item in various domains [38].

III. MAIN CONTRIBUTIONS

This article aims to forge a nexus between the realms of deep learning and transfer learning, harnessing principles akin to human learning to give rise to the framework known as Deep Transfer Learning (DTL) [39]. The study explores various cutting-edge deep learning techniques based on self-learned patterns, leading to the development of deep learning models. The focus is on investigating how human-like learning processes can inspire learning algorithms for transferring information from one scenario to another [40]. The potential causes of transfer learning settings, including distributions, posterior probabilities, learning functions, and classification tasks, are intriguing. The study also examines different deep neural network (DNN) based strategies for creating enhanced synthetic datasets to improve object detection performance. The DTL was developed by transferring learning from a previously trained model to a new model created based on the information of the first model using a dataset distinct from the dataset used to build the pre-trained model, thereby overcoming the limitations of DL [41].

Certainly, our research introduces several distinct facets that contribute to its novelty and potential advantages over existing state-of-the-art techniques in the domain of object detection and transfer learning within smart city surveillance systems. Enhanced object detection accuracy, our research harnesses the Single Shot MultiBox Detector (SSMD) architecture, specifically SSD-512 and SSD-300 models, trained with the COCO dataset. The meticulous incorporation of transfer learning (TL) techniques empowers these models to offer significantly improved object detection accuracy compared to conventional methods. Synthetic dataset augmentation, we explore the utilization of synthetic datasets constructed from the KITTI dataset, enhancing the richness and diversity of the training data. By introducing variations such as lighting conditions and object orientations, our method enhances the model's adaptability to real-world scenarios, setting it apart from traditional training approaches. Proposed research strategically employs transfer learning by fine-tuning pre-trained models from the COCO dataset, SSD-512 and SSD-300 models. This approach significantly reduces the requirement for labeled examples and enhances training efficiency, showcasing the innovative adaptation of existing knowledge to improve object detection in smart city surveillance. Comprehensive validation and performance metrics, our study offers a thorough evaluation using various performance metrics, such as accuracy, precision, recall, and F1-score. This extensive analysis ensures a robust assessment of the proposed models' efficiency and provides a clear comparison with established methodologies, underlining the strength and reliability of our proposed techniques.

These highlighted facets collectively contribute to our research's uniqueness and potential advantages, underscoring its significance in advancing the realm of object detection within smart city surveillance and demonstrating its superiority over current methodologies.

IV. DEEP LEARNING

Within machine learning (ML), an intriguing new avenue known as deep learning (DL) has emerged. DL has demonstrated remarkable promise, particularly in the domain of computer vision. This methodology involves training multi-layer artificial neural networks (ANNs) utilizing diverse DL techniques, enabling the examination of various features through a plethora of DL systems and models [42]. The effectiveness of these techniques can be enhanced through iterative processes, facilitating the discernment of intricate data-driven patterns. In the symbolic learning paradigm of DL systems, each layer receives input from its predecessor and passes it forward, with the initial layer focusing on fundamental and coarse features. Deeper network layers acquire a nuanced understanding, extracting more precise features than the raw datasets, thereby shaping distinctive characteristic traits that guide the progressive evolution of AI software [43]. Due to its competence in representing intricate data like images and sounds, DL has found widespread adoption across diverse industries, making it a focal point within ML. Notably, DL encompasses four key categories of techniques, namely deep unsupervised learning (DUSL), deep supervised learning (DSL), deep reinforcement learning (DRL), and deep semi-supervised learning (DSSL), each leveraging labeled datasets [44]. DSL, in particular, includes deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs).

Recently, the CNN approach, as a DL technique, has exhibited remarkable prowess in image recognition. It encompasses an array of parameters: layer count, bias, neuron count, activation functions, weights, stride, learning rate, filter dimensions, and more. The architecture of CNN can be categorized into seven distinct types: attention mechanisms, feature map exploitation, multi-path structures, breadth and depth adjustments, channel enhancements, and spatial exploration. This architecture is divided into two fundamental components: feature extractors and classifiers. The CNN extractor includes a stack of convolution layers and a max-pooling layer, while the CNN classifier comprises fully connected and softmax layers at the final stage. Prominent

spatial exploration CNN architectural models encompass Alexnet, VGG, GoogleNet, LeNet, and ResNet. ResNet and GoogleNet are specifically tailored for large-scale data processing, while VGG represents a more general architecture [45]. It is worth noting that DL systems often demand a greater volume of data and training iterations than conventional ML systems to attain optimal results.

V. DATASET

Computer vision is an innovative technology that empowers computers to understand and interpret images. By providing the correct image datasets, data scientists can teach computers to operate as if they had their own eyes. This technology lays the groundwork for various groundbreaking discoveries and advancements, such as facial recognition and self-driving cars. In computer vision, a dataset is a thoughtfully curated collection of digital images used by developers to assess, train, and test the performance of their algorithms. If we aim to identify items in photos with bounding boxes, we require an object detection dataset. This dataset includes both images (or videos) and annotations. The KITTI dataset provides a complete set of visual tasks organized through an autonomous driving platform [46]. The comprehensive benchmark comprises a multitude of tasks, encompassing not only stereo and optical flow but also extending its scope to include monocular images and bounding boxes derived from object detection datasets. Within this framework, bounding boxes were meticulously incorporated into a substantial set of 7481 images.

The approach taken in this research regarding augmented synthetic datasets is indeed an intriguing facet of the study. Augmented synthetic datasets offer the advantage of diversifying the training data, which is critical for the robustness and generalization of deep neural networks. The construction of augmented synthetic datasets involves the generation of additional data images through data augmentation, which includes rotation and lighting conditions. Moreover, by splitting the data into training and validation sets, validation protocols are crucial for assessing the performance of the model trained on these synthetic datasets. Furthermore, the discussion encompasses details on the evaluation metrics employed to validate the effectiveness of the augmented synthetic datasets. The metrics that we have used to measure the accuracy, precision, recall, and F1-score, encountered during the construction and validation processes. It's important to the construction and validation of augmented synthetic datasets are imperative to provide a deeper insight into the methodologies employed, the quality of the generated data, and the overall impact on the performance of the deep learning models for object detection within the smart city surveillance context. The original dataset is available at Kaggle website, https://www.kaggle.com/datasets/klemenko/kitti-dataset

VI. CNN ARCHITECTURES

In Fig. 1, we can observe the fundamental architecture of CNN [47], which comprises three distinct layers: input, hidden (latent), and output. The hidden (latent) layer is further categorized into fully connected, pooling, or convolutional layers, each serving a unique function in the network's operations.

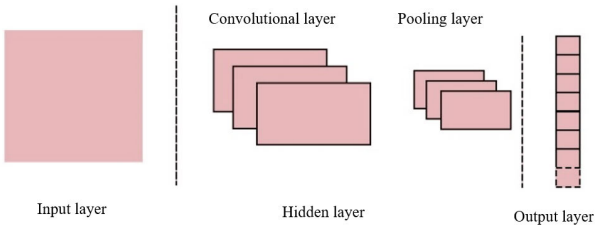

FIGURE 1. An illustrative diagram depicting the fundamental architecture of a Convolutional Neural Network (CNN).

A. THE CONVOLUTIONAL LAYERS

Figure 2 illustrates the structure of a traditional, discrete convolution layer. This layer occupies the highest position in a Convolutional Neural Network (CNN) architecture. It operates iteratively, employing convolution processes to generate a dynamic output function from the provided functions [48]. Within this convolutional layer, the filters or feature maps consist of a multitude of neurons, typically sized proportionally to the input data. Consequently, evaluating the convolution of individual receptors unveils the responsiveness of these neurons. This evaluation relies on the cumulative weight of input neurons and the application of the activation function.

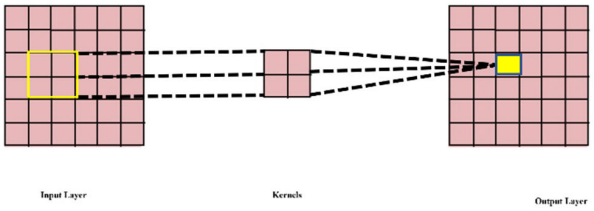

FIGURE 2. The architectural blueprint of a discrete convolutional layer.

B. MAX-POOLING LAYER

Figure 3 illustrates the intricate process of constructing maximum combining layers. This involves the generation of multiple interlaced structures originating from the output of segmented convolution layers. To initiate this process, the grid's most elevated values were initially organized into a matrix. Following this, the operator meticulously computed each matrix, discerning whether to derive its average or select its maximum value.

C. FULLY CONNECTED LAYERS

Figures 4 and 5 elucidate the distinct configurations of two critical components within our framework: a conventional fully connected layer and a comprehensive Convolutional Neural Network (CNN) encompassing all three layers. The

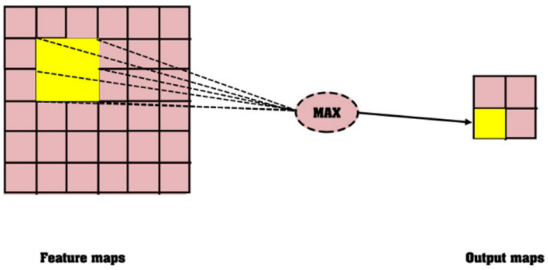

FIGURE 3. The architectural composition of max-pooling layers entails a sophisticated configuration.

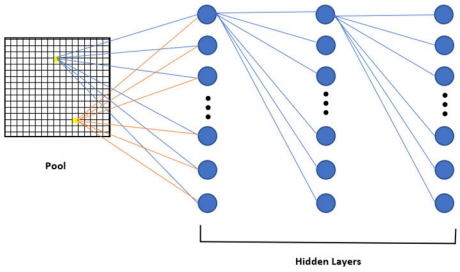

FIGURE 4. The complete configuration of the fully connected layers.

nomenclature 'fully connected layer' signifies a CNN in which nearly 90% of its structural elements are engaged. This layer orchestrates the seamless exchange of input data with networks characterized by predetermined vector lengths [47]. At this juncture, the input dataset undergoes a transformative metamorphosis before its eventual classification. Moreover, while conserving the fidelity of information, the convolutional layer undergoes a similar process of refinement. Ultimately, neurons from each antecedent layer synergize to form the fully connected layer, a pivotal element employed in the intricate task of image classification.

It's important to recognize that the CNN architecture presented above may not be inherently optimized for tackling the complexities of computer vision challenges, as its design predominantly emphasizes object recognition. Consequently, the development of a network configuration finely tuned to the specific problem domain assumes paramount importance in the pursuit of attaining peak performance. Nevertheless, empirical evidence underscores that the CNN we have constructed exhibits the capability to deliver the sought-after solutions.

VII. SINGLE-SHOT MULTI-BOX DETECTOR ARCHITECTURE

The SSMD model, introduced in reference [49], employs network-wide maps to predict object locations and localization scores simultaneously, accommodating standard boxes with varying image attributes. This approach diverges from earlier methods by utilizing multiple maps throughout the network to enhance detection speed. SSMD, leveraging Convolutional Neural Networks (CNNs), directly generates bounding boxes based on object probabilities for each class. Subsequently, it employs a non-maximum suppression (NMS) technique to derive the ultimate detection outcomes. The SSMD model comprises two prevalent architectures: SSD-300 and SSD-512, designed for input sizes of 300 and 512, respectively [50]. SSD-300, illustrated in Figure 1, utilizes VGG16_15 as the base network, featuring 512 38 $\times$ 38-pixel feature maps. The subsequent part of the model incorporates convolution layers to amalgamate multi-layer features, generating bounding boxes for each class. Following VGG16_15, five convolution layers are deployed, yielding feature maps of sizes (19×19), (10×10), (5×5), (3×3), and (1×1), with dimensions of 1024, 1024, 512, 256, and 256, respectively. SSD-300 combines six distinct feature maps to generate the desired bounding box. It's worth noting that while the aforementioned CNN structure is primarily tailored for object recognition, custom network configurations are imperative for optimizing performance in specific problem domains [51]. Nevertheless, empirical results underscore the capability of this constructed CNN to deliver satisfactory solutions.

VIII. RGB COLOR SPACE

In the realm of color images, various hues are harmoniously blended to create a cohesive visual representation. In the context of our investigation, these representations serve as repositories for the precise delineation of colors, specifying both the quantities and varieties of color channels, which are referred to as the color space. Notably, we focused on the RGB (Red, Green, Blue) color space, often recognized as the quintessential color model for digital images. In this model, RGB images are envisaged as 3-D arrays, where each dimension corresponds to one of the three primary colors: red (R), green (G), and blue (B). RGB is widely embraced as the go-to color space for digital images due to its seamless alignment with the fundamental principles of color mixing, making it the ideal choice for rendering images on monitors and screens. To illustrate this, Figure 6 visually depicts the RGB color channels of a true-color image [52]. In essence, genuine image colors emerge from the skillful interplay of different hues residing within the RGB channels.

IX. TRANSFER LEARNING

Initiating a deep learning (DL) model from scratch is atypical due to the substantial data and time investment required for it to converge effectively. Consequently, pre-trained models often come into play, serving as a foundation or feature extractor. Transfer learning (TL), a method in which knowledge gleaned from one domain is transposed to another through the reuse of a pre-trained model, becomes instrumental in this context. TL streamlines data requirements, augments training efficiency, and bolsters accuracy. In this study, we harnessed TL by leveraging pre-trained models derived from the Single Shot MultiBox Detector (SSD) trained on the COCO dataset, addressing the scarcity of labeled examples. We adopted SSD512 as

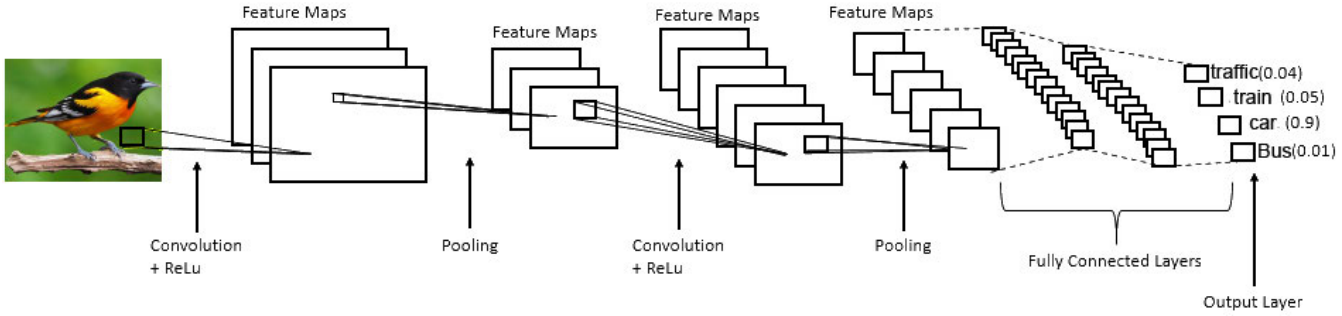

FIGURE 5. The configuration of a conventional complete Convolutional Neural Network (CNN).

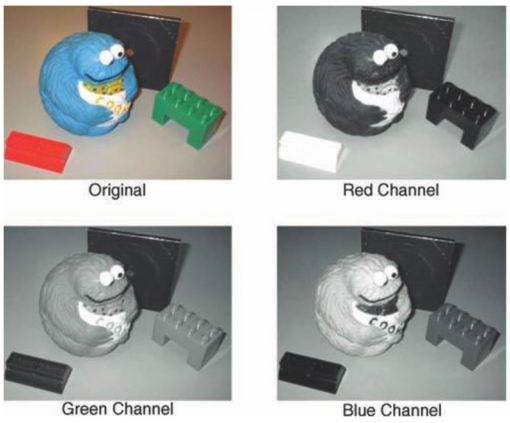

FIGURE 6. The RGB channels.

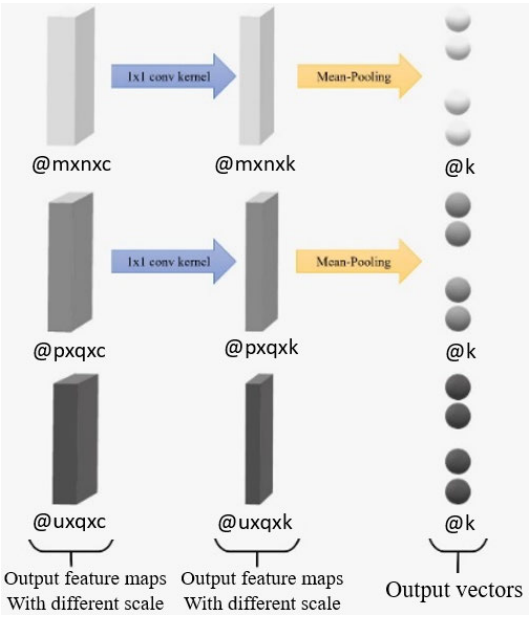

FIGURE 7. Working of fully convolutional structure.

the base model, characterized by the VGGNet-16 network structure and the Multi-Scale Convolution Feature Detection Network (MSCFDN) for feature extraction and bounding box creation. Fine-tuning, a pivotal TL technique, was employed to adapt model parameters for class predictions and bounding boxes based on the input dataset. During model training, batch size, weight decay, momentum, iteration count, and base learning rate were meticulously configured at 64, 0.05, 0.9, 50,000, and 0.01 divided by 10, respectively. At 12,000 and 20,000 iterations, minimal compromises were observed due to deploying a single deep neural detection network.

The holistic convolution architecture obviates the need for consistency considerations. Fig. 7 elucidates the operational framework of the complete convolution structure, which initially processes input feature maps of diverse shapes by applying a 1×1 convolution kernel within the same channel. The resultant feature map maintains identical channels for all classes. Subsequently, a mean-pooling layer generates an average vector with class-appropriate dimensions for each channel. In this manner, the CNN leverages the full convolution architecture to adeptly address complex inquiries.

Figure 8 provides an insightful depiction of the comprehensive architecture of VGG16. This intricate structure is visually represented, with convolution layers in blue, maximum-pooling layers in red, and fully connected layers in yellow. Notably, VGG16 operates with input images of dimensions $224 \times 224 \times 3$. VGG16 has garnered acclaim for its exceptional generative prowess. In a noteworthy departure from its predecessor, Alexnet, VGG16 has achieved a significant reduction in the number of parameters, a feat accomplished through the adoption of smaller convolution kernels measuring $3 * 3$. A discerning observation of Figure 10 reveals that the final three layers of VGG16 are comprised of 4096 fully connected elements.

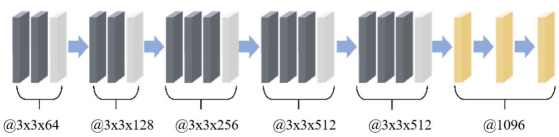

FIGURE 8. The VGG16 neural network structure.

X. THE FULLY CONVOLUTIONAL ARCHITECTURE AND DETAILED ANALYSIS OF VGG16

We adopted a fully convolutional architecture to accommodate the varying input sizes stemming from a spectrum of weld defects. Given the substantial diversity in the sizes of these defects, they necessitated cropping into distinct dimensions. Traditional fully connected layers function in accordance with the mechanism illustrated by Eq. 5, wherein the outcome of the product between W and X governs the constitution of the initial fully connected layer. This layer is subsequently supplied with the ultimate feature map derived from the concluding convolutional layer, thereby preserving the continuity of the preceding feature map dimensions.

$$O = h(Wx + b) \tag{1}$$

XI. PERFORMANCE EVALUATION

In assessing the model's efficacy, a range of metrics were meticulously employed, encompassing accuracy, positive predictive value (PPV), sensitivity (also known as recall), and the F1 score. The F1 score, being a harmonious amalgamation of precision and recall, was the gauge for evaluating the system's prowess in detection. Table 1 was used to systematically tabulate the instances of true positives (T.P.), true negatives (T.N.), false positives (F.P.), false negatives (F.N.), and instances where nothing was detected (N.P.), thus offering a comprehensive representation of the outcomes [53].

TABLE 1. The accuracy for evaluation of object detection.

Detection	Position
T.P.	Both the identification of geographical locations and the categorization of object types are executed
F.P.	The discrepancies observed can be attributed to one of two scenarios: firstly, the object type may be accurate, yet its placement is inaccurate; secondly, the object type itself may be correct, but its positioning is erroneous.
T.N.	The count of genuine images accurately categorized as authentic images.
F.N	The count of images that have been inaccurately classified.

Calculation of Diverse Metrics Utilizing the Following Formulas:

$$\text{Acc} = \frac{(\text{T.P.} + \text{T.N.})}{(\text{T.P.} + \text{F.P.} + \text{F.N.} + \text{T.N.})} \tag{2}$$

$$\text{P} = \frac{T.P.}{\text{T.P.} + \text{F.P.}} \tag{3}$$

$$\text{Recall} = \frac{T.P.}{\text{N.P.}} \tag{4}$$

$$\text{F score} = \frac{2 \times (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \tag{5}$$

XII. THE PROPOSED METHOD

During this investigation, we harnessed the KITTI dataset, an extensive collection encompassing a total of 7481 images. This dataset underwent an initial bifurcation into two distinct subsets, with 80% allocated to the training set and the remaining 20% earmarked for rigorous testing. To facilitate the forthcoming transfer learning endeavor, a meticulous preprocessing step was executed. Specifically, all images were meticulously resized, with dimensions set at 300 × 300 pixels for the training of SSD 300 and a larger 512 × 512 pixel format for the training of SSD 512, ensuring optimal readiness for subsequent model development and assessment. Subsequently, the testing set was used to appraise the performance of the proposed technique. Accuracy, precision, recall and F1-score were then computed, as illustrated in Figure 9.

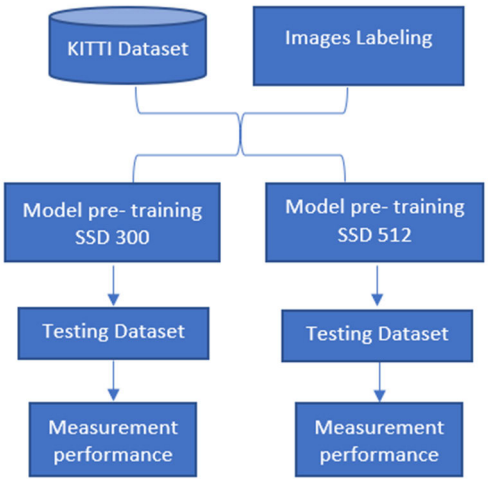

FIGURE 9. The main framework designs.

An object detection method that utilizes SSMD to identify object type and position in a digital image was proposed. A correct label was created to retrain the SSD KITTI, and labeling was employed to generate labels for the training data. The training data was composed of intermediate resolution samples with dimensions of 512 × 512. To ensure compatibility with the TensorFlow framework, data preprocessing involved transforming the obtained KITTI label and image. The dataset underwent a process of retraining within the SSMD, where novel weight configurations were employed to facilitate object detection within the digital image. Python-based Keras served as the deep learning library, while TensorFlow, developed and powered by Google Inc., functioned as the backend for object recognition. A visual representation of the suggested approach for object detection is illustrated in Figure 10.

The system under development trained an appropriate label for the original image and securely stored it in the SSMD. Subsequently, upon completing this training phase, a new set of weights was acquired. These weights were then employed to determine the optimal weight configurations necessary for object detection. The selection of these weight configurations was guided by the transformation of the loss coefficients obtained during the Transfer Learning (TL)

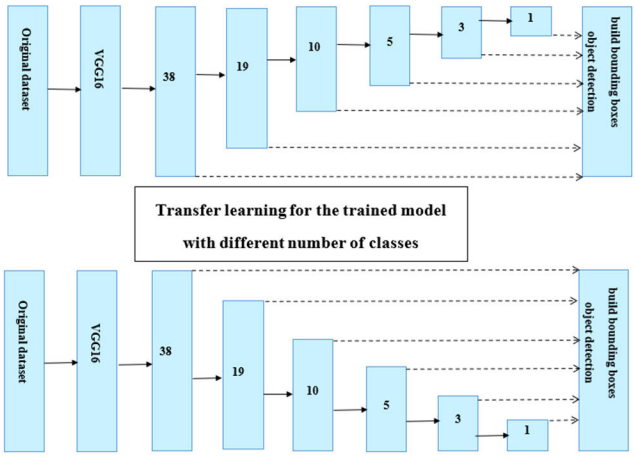

FIGURE 10. Transformation learning (TL) analysis of SSD-300: extracted feature maps from five convolution layers following VGG16 layer.

process. The primary objective here was to establish a meticulously documented implementation protocol for the SSMD. This protocol aimed to simplify the process for users, enabling them to utilize the code efficiently and either build or apply a model. This approach was favored over alternative implementation schemes, such as Keras, which often provide minimal information or elucidation in their tutorials and instructions.

The methodology adopted in this research embarks on a meticulous approach encompassing the architecture nuances and transfer protocols, particularly focusing on the intricacies of the network design and the transfer learning protocols. The study predominantly revolves around the Single Shot MultiBox Detector (SSMD) architecture, an innovative framework designed for object detection. Within the SSMD, the underlying architecture comprises both SSD-300 and SSD-512 models, each tailored for specific input image dimensions. SSD-300, employing the VGG16_15 architecture as its base network, operates with multiple convolutional layers generating feature maps of varying dimensions, followed by subsequent convolution layers for multi-layer feature integration. Similarly, SSD-512, designed for larger input images, demonstrates an enhanced ability to detect and delineate objects with greater precision.

Moreover, transfer learning (TL) is a pivotal component employed in this research. TL involves the transference of learned information from pre-trained models, enhancing the efficiency of training models on new datasets. The study adopts TL by utilizing pre-trained models derived from the COCO dataset for the SSD-512 and SSD-300 base models. This technique mitigates the scarcity of labeled data by fine-tuning the model parameters based on the specific dataset being used for training. The protocol of the TL process involves meticulous configuration of various parameters such as batch size, weight decay, momentum, iteration count, and base learning rate to effectively adapt the model parameters for the specific classification and bounding box predictions on the new dataset. Additionally, the utilization of TensorFlow and Keras within the Python framework serves as the bedrock for developing and implementing the object detection system, ensuring a robust and efficient model that outperforms conventional methodologies.

Furthermore, the study also elaborates on the pre-processing steps on the KITTI dataset, which involves resizing the images to meet the required dimensions for SSD training, ensuring optimal readiness for subsequent model development and evaluation. The efficacy of the proposed model was rigorously tested and evaluated through a range of performance metrics, emphasizing accuracy, precision, recall, and F1-score. This comprehensive assessment included a thorough breakdown of the computer systems and software utilized for the object detection process through deep learning, along with the details of the testing procedures. The meticulous evaluation, extensive analysis, and comparison with previous techniques showcased the superiority of the proposed SSMD-512 and SSMD-300 models in terms of accuracy and precision for object detection, thereby emphasizing their potential for advancing the precision of object detection beyond current state-of-the-art studies.

In essence, the research's methodology intricately involves the SSMD architecture's detailed specifications and the strategic implementation of transfer learning, underscoring the pivotal role of these components in enhancing object detection accuracy within smart city surveillance systems.

XIII. RESULTS AND DISCUSSION

Figure 11 serves as a visual testament to the origins of SSD300 and SSD512, affirming the efficacy of our proposed method in precisely delineating object positions. The accuracy assessment was conducted through a comprehensive evaluation encompassing multiple criteria, meticulously detailed in Table 1. In Table 2, we provide an exhaustive breakdown of the computer system and software employed in the automated object detection process via deep learning, as well as a concise overview of the testing procedure.

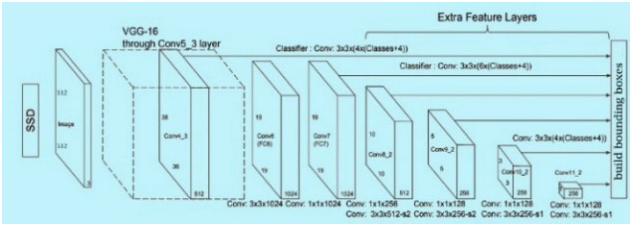

FIGURE 11. The origins of SSD300 and SSD512 can be traced to models designed with input image dimensions set at 512 pixels in both width and height.

Table 3 furnishes invaluable insights into the accuracy, F1-score, and computational runtime for the tested images.

TABLE 2. Context of transfer learning and training.

Details	Hardware devices
Windows 10	OS
Intel(R) Core (TM) i7-10750 H; CPU @ 2.60GHz & 2.59 GHz	CPU
GeForce RTX 7020, 8 GB	GPU
32 GB	Main memory
Python 3.7	Language

TABLE 3. Classification of datasets with evaluation metrics and time required for each epoch's execution.

Dataset	Class	Model	P	R	FI	ACU	Run time/ms
KITTI	8	SSD-512-TL	97.67	97.34	97.50	99.89	46
KITTI	8	SSD-300-TL	97.88	98.82	98.34	98.74	31

Additionally, Table 4 presents a dataset subjected to scrutiny through the Transfer Learning (TL) model, specializing in object detection and localization, achieved through bounding boxes. This study harnessed two distinct deep learning models, partitioning 80% of the learning dataset for rigorous training and reserving the remaining 20% for meticulous testing, sourced from the diverse KITTI dataset, encompassing a total of eight object classes.

Numerous deep learning algorithms have been proposed for feature extraction and object detection. However, the accuracy of object detection remains a challenge. Deep learning approaches have surpassed conventional machine learning methods in visual object segmentation, and their feature extraction efficiency may be scaled up based on processing power, model complexity, and training data quantity. This study's analysis indicates that the proposed deep learning method outperforms typical deep learning techniques in terms of performance measures, as demonstrated in Table 5, suggesting its potential for improving object detection precision beyond current state-of-the-art studies.

Certainly, while our approach exhibits notable strengths, it also encompasses inherent limitations and areas for potential improvement. One significant limitation lies in the need for more comprehensive discussions or explorations regarding the model's performance in challenging scenarios, especially when collecting high-quality datasets or complex background settings. As our methodology heavily relies on transfer learning and synthetic dataset augmentation, the quality and representativeness of these synthesized data become crucial factors impacting the model's real-world applicability. Additionally, the applicability of the proposed system might encounter challenges when handling real-time processing in smart city surveillance due to the computational load associated with deep neural networks, an aspect necessitating further optimization to ensure real-time implementation feasibility. Improvement opportunities exist in refining the synthetic data generation process to simulate a wider range of real-world scenarios better, enhancing the robustness and adaptability of the models. Furthermore, the generalizability of our method across various domains within smart city applications would greatly benefit from an extensive evaluation of diverse

TABLE 4. Conducting a testing procedure on a sample from the KITTI dataset using a model that has undergone Transfer Learning (TL).

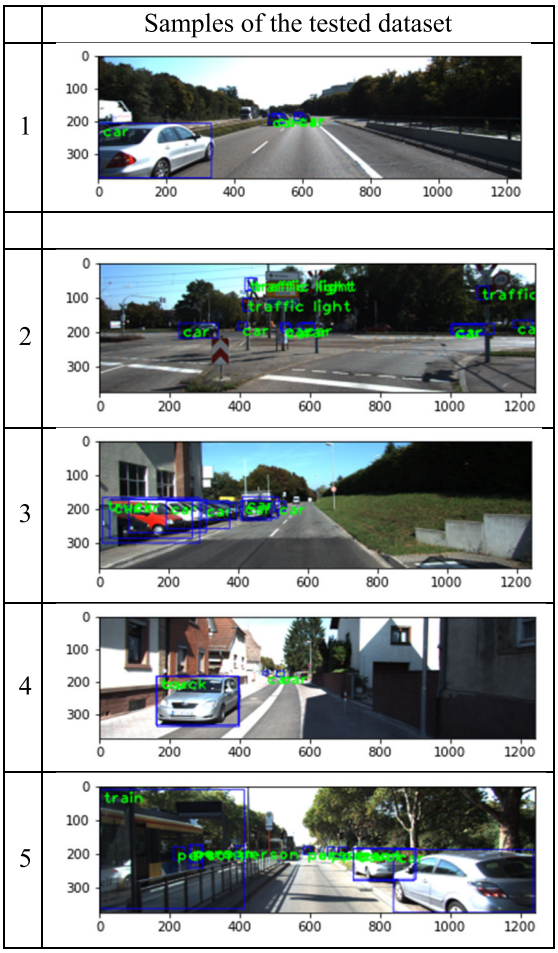

TABLE 5. The comparison for the proposed method with achievements of previous techniques.

Author	Algorithms	precision	accuracy
Olivier Janssens et al (2017) [54]	CNN	95%	
Christopher Dahlin Rodin et al.(2018) [55]	GMM and CNN	-	92.5%
Yunyoung Nam et al.(2018) [56]	regions of interest (ROIs)	-	92.7%
Aparna et al.(2019) [16]	CNN	-	97.08%
Proposed method	**SSD-512-TL**	**97.67**	**99.89**
Proposed method	**SSD-300-TL**	**97.88**	**98.74**

datasets and practical implementation scenarios, which will underscore its reliability and effectiveness in real-world deployments. Identifying and addressing these limitations will substantially contribute to the manuscript's credibility and pave the way for more practical and impactful real-world applications of the proposed approach.

XIV. CONCLUSION

This study introduces an exceptionally efficient object detection system that seamlessly integrates SSMD (Semantic Segmentation and Object Detection) and Transfer Learning (TL). The initial phase involved constructing a model through pre-training to facilitate transfer learning. Subsequently, this model was employed for the pre-training of SSMD-512 and SSMD-300, utilizing the KITTI dataset. A comprehensive evaluation of the model's performance was conducted using a set of rigorous performance metrics. The experimental results clearly demonstrated a significant enhancement in the object detection accuracy of the SSMD system.

Nevertheless, it was discerned that certain minor elements, such as edges and corners, required careful consideration within the current SSMD to mitigate false alarms. This object detection methodology holds the potential to automate error-prone and time-intensive tasks, precisely identifying object locations and bounding boxes. The integration of Transfer Learning played a pivotal role in detecting items and pinpointing objects across diverse datasets, catering to the needs of AI scientists. The remarkable performance of this novel approach can be attributed to the symbiosis of Deep Learning (DL) and Transfer Learning, allowing for the automatic extraction of image features, while traditional Machine Learning (ML) necessitates manual feature selection. Furthermore, DL's data-hungry nature allows Transfer Learning to be flexibly applied across various related scenarios, facilitating knowledge transfer. This system can be leveraged for the automatic categorization, deduplication, and organized storage of digital photographs upon the completion of object position detection. When juxtaposed with existing methods like CNNs and conventional SSMDs, the newly developed Transfer Learning model, SSMD-512, and SSMD-300 exhibit peak performance, ensuring heightened accuracy and efficiency.

ACKNOWLEDGMENT

The author Safa Riyadh Waheed wishes to convey his profound gratitude to those whose contributions have enriched this research endeavor. Foremost, he wishes to offer his heartfelt appreciation to his esteemed supervisors, whose invaluable insights, guidance, and unwavering support have been instrumental in the course of this study. The authors would also like to thank Prince Sultan University, Riyadh, Saudi Arabia, for its support.

CONFLICT OF INTEREST

The authors declare no conflict for this research.

FUNDING

No funding was received for this research.

REFERENCES

[1] S. Mujeeb, T. A. Alghamdi, S. Ullah, A. Fatima, N. Javaid, and T. Saba, "Exploiting deep learning for wind power forecasting based on big data analytics," *Appl. Sci.*, vol. 9, no. 20, p. 4417, Oct. 2019.

[2] M. A. Khan, T. Akram, M. Sharif, K. Javed, M. Raza, and T. Saba, "An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection," *Multimedia Tools Appl.*, vol. 79, nos. 25–26, pp. 18627–18656, Jul. 2020.

[3] T. Saba and A. Rehman, "Effects of artificially intelligent tools on pattern recognition," *Int. J. Mach. Learn. Cybern.*, vol. 4, no. 2, pp. 155–162, Apr. 2013.

[4] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[5] T. Saba, M. A. Khan, A. Rehman, and S. L. Marie-Sainte, "Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction," *J. Med. Syst.*, vol. 43, no. 9, p. 289, Sep. 2019.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[7] S. R. Waheed, N. M. Suaib, M. S. M. Rahim, M. M. Adnan, and A. A. Salim, "Deep learning algorithms-based object detection and localization revisited," *J. Phys., Conf.*, vol. 1892, no. 1, Apr. 2021, Art. no. 012001.

[8] X. Peng, B. Sun, K. Ali, and K. Saenko, "Learning deep object detectors from 3D models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1278–1286.

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[10] G. Georgakis, A. Mousavian, A. C. Berg, and J. Kosecka, "Synthesizing training data for object detection in indoor scenes," 2017, *arXiv:1702.07836*.

[11] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3296–3297.

[12] P. S. Rajpura, H. Bojinov, and R. S. Hegde, "Object detection using deep CNNs trained on synthetic images," 2017, *arXiv:1706.06782*.

[13] P. Rajpura, A. Aggarwal, M. Goyal, S. Gupta, J. Talukdar, H. Bojinov, and R. Hegde, "Transfer learning by finetuning pretrained CNNs entirely with synthetic images," in *Proc. Nat. Conf. Comput. Vis., Pattern Recognit.* Cham, Switzerland: Springer, 2017, pp. 517–528.

[14] J. Talukdar, S. Gupta, P. S. Rajpura, and R. S. Hegde, "Transfer learning for object detection using state-of-the-art deep neural networks," in *Proc. 5th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2018, pp. 78–83.

[15] D. Zhang, J. Hu, F. Li, X. Ding, A. K. Sangaiah, and V. S. Sheng, "Small object detection via precise region-based fully convolutional networks," *Comput., Mater. Continua*, vol. 69, no. 2, pp. 1503–1517, 2021.

[16] J. Wang, Y. Wu, S. He, P. K. Sharma, X. Yu, O. Alfarraj, and A. Tolba, "Lightweight single image super-resolution convolution neural network in portable device," *KSII Trans. Internet Inf. Syst. (TIIS)*, vol. 15, no. 11, pp. 4065–4083, 2021.

[17] J. Wang, Y. Zou, P. Lei, R. S. Sherratt, and L. Wang, "Research on recurrent neural network based crack opening prediction of concrete dam," *J. Internet Technol.*, vol. 21, no. 4, pp. 1161–1169, 2020.

[18] J. Zhang, J. Sun, J. Wang, and X.-G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 8, pp. 8427–8440, Aug. 2021.

[19] S. He, Z. Li, Y. Tang, Z. Liao, F. Li, and S.-J. Lim, "Parameters compressing in deep learning," *Comput., Mater. Continua*, vol. 62, no. 1, pp. 321–336, 2020.

[20] S. Zhou and B. Tan, "Electrocardiogram soft computing using hybrid deep learning CNN-ELM," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105778.

[21] S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 393–400, Feb. 2019.

[22] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[23] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.

[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[27] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2155–2162.

[28] W. Wang, H. Liu, J. Li, H. Nie, and X. Wang, "Using CFW-Net deep learning models for X-ray images to detect COVID-19 patients," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 199–207, 2020.

[29] W. Wei, J. Yongbin, L. Yanhong, L. Ji, W. Xin, and Z. Tong, "An advanced deep residual dense network (DRDN) approach for image super-resolution," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 2, pp. 1592–1601, 2019.

[30] C. Szegedy, S. Reed, D. Erhan, D. Anguelov, and S. Ioffe, "Scalable, high-quality object detection," 2014, *arXiv:1412.1441*.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

[32] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[34] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[35] K. A. Kadhim, F. Mohamed, Z. N. Khudhair, and M. H. Alkawaz, "Classification and predictive diagnosis earlier Alzheimer's disease using MRI brain images," in *Proc. IEEE Conf. Big Data Analytics (ICBDA)*, Nov. 2020, pp. 45–50.

[36] W. Wang, Y. Yang, J. Li, Y. Hu, Y. Luo, and X. Wang, "Woodland labeling in Chenzhou, China, via deep learning approach," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 1393–1403, 2020.

[37] W. Sun, G. Zhang, X. Zhang, X. Zhang, and N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tool. Appl.*, vol. 80, no. 20, pp. 30803–30816, 2021.

[38] X. Zhang, X. Sun, W. Sun, T. Xu, P. Wang, and S. Kumar Jha, "Deformation expression of soft tissue based on BP neural network," *Intell. Autom. Soft Comput.*, vol. 32, no. 2, pp. 1041–1053, 2022.

[39] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[40] T. Zutalin. *README*. Accessed: Apr. 8, 2021. [Online]. Available: https://github.com/tzutalin/labelImg/blob/master

[41] Y. Xu, M. Zhu, S. Li, H. Feng, S. Ma, and J. Che, "End-to-end airport detection in remote sensing images combining cascade region proposal networks and multi-threshold detection networks," *Remote Sens.*, vol. 10, no. 10, p. 1516, Sep. 2018.

[42] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and M. Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, pp. 195–204, Feb. 2019.

[43] E. Davies and M. Turk, *Advanced Methods and Deep Learning in Computer Vision*. Amsterdam, The Netherlands: Elsevier, 2021.

[44] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using recurrent neural networks (RNN)," in *Proc. Integr. Commun. Navig. Surveill. (ICNS)*, 2016, pp. 1–5.

[45] F. H. Najjar, H. M. Al-Jawahry, M. S. Al-Khaffaf, and A. T. Al-Hasani, "A novel hybrid feature extraction method using LTP, TFCM, and GLCM," *J. Phys., Conf.*, vol. 1892, no. 1, Apr. 2021, Art. no. 012018.

[46] A. Santoro, R. Faulkner, D. Raposo, J. Rae, M. Chrzanowski, T. Weber, D. Wierstra, O. Vinyals, R. Pascanu, and T. Lillicrap, "Relational recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[47] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, Mar. 2021.

[48] K. Liu, G. Kang, N. Zhang, and B. Hou, "Breast cancer classification based on fully-connected layer first convolutional neural networks," *IEEE Access*, vol. 6, pp. 23722–23732, 2018.

[49] A. Kumar and S. Srivastava, "Object detection system based on convolution neural networks using single shot multi-box detector," *Proc. Comput. Sci.*, vol. 171, pp. 2610–2617, Jan. 2020.

[50] Z. Chen, T. Zhang, and C. Ouyang, "End-to-end airplane detection using transfer learning in remote sensing images," *Remote Sens.*, vol. 10, no. 1, p. 139, Jan. 2018.

[51] H. Zhao, F. Liu, H. Zhang, and Z. Liang, "Convolutional neural network based heterogeneous transfer learning for remote-sensing scene classification," *Int. J. Remote Sens.*, vol. 40, no. 22, pp. 8506–8527, Nov. 2019.

[52] Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using Sentinel-1 SAR images," *Remote Sens. Lett.*, vol. 9, no. 8, pp. 780–788, Aug. 2018.

[53] S. R. Waheed, M. S. M. Rahim, N. M. Suaib, and A. A. Salim, "CNN deep learning-based image to vector depiction," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 20283–20302, May 2023.

[54] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 151–159, Feb. 2018.

[55] C. D. Rodin, L. N. de Lima, F. A. D. A. Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold, "Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[56] Y. Nam and Y.-C. Nam, "Vehicle classification based on images from visible light and thermal cameras," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–9, Dec. 2018.

SAFA RIYADH WAHEED was born in Iraq. He received the master's degree in computer science from Universiti Teknologi Malaysia (UTM), Malaysia. He is currently a Computer Science Lecturer with the Department of Computer Technology Engineering, UTM. He is also a computer science teacher, an author, and a researcher. His work spans several sub-fields of computer science, including computer vision, digital image processing, medical image processing, data and image security, computer networking, real-time applications, and the public understanding of computer science.

NORHAIDA MOHD SUAIB is currently a Lecturer with the Faculty of Computing, Universiti Teknologi Malaysia (UTM), where she teaches many computer graphics-related subjects at the undergraduate level and supervised many postgraduate research. She is also a Research Fellow with the UTM Big Data Center (UTM BDC), Ibnu Sina Institute for Scientific and Industrial Research (ISI-SIR), UTM. She is an active member of the UTM ViCubeLab Research Group, a research group dedicated to virtual (virtual reality/virtual environment), visualization, and vision. She specializes in the field of computer graphics, particularly interactive computer graphics, collision detection, physics-based models, visualization, and virtual reality. Her current research interests include computer graphics and computer games technology in the preservation of both tangible and intangible cultural heritage; where modeling, object/action recognition, and reconstruction play important parts to increase the realism of computer-generated scenes.

MOHD SHAFRY MOHD RAHIM received the B.Sc. degree (Hons.) in computer science and the M.Sc. degree in computer science from Universiti Teknologi Malaysia (UTM), in 1999 and 2002, respectively, and the Ph.D. degree in spatial modeling from Universiti Putra Malaysia, in 2008. He is currently a Professor of image processing with the School of Computing, UTM. He is also the Deputy Director of the Centre for Joint Programme, UTMSPACE. He focused his research together with the UTM ViCubeLab Research Group, Faculty of Computing, UTM. He is also an expert in the research area of computer graphics and image processing.

AMJAD REHMAN KHAN (Senior Member, IEEE) received the Ph.D. degree from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2010, specializing in information security using image processing techniques. He was a recipient of the Rector Award for the 2010 Best Student from UTM Malaysia. He is currently an Associate Professor with the College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several projects and completed projects funded by MoHE Malaysia, Saudi Arabia. His research interests include bioinformatics, the IoT, information security, and pattern recognition.

SAEED ALI BAHAJ received the Ph.D. degree from Pune University, India, in 2006. He is currently an Associate Professor with the Department of Computer Engineering, Hadramout University, Yemen, and the Department of Management Information Systems, College of Business Administration (COBA), Prince Sattam bin Abdulaziz University. His research interests include artificial intelligence, information management, forecasting, information engineering, big data, and information security.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently a Full Professor with the College of Computer and Information Sciences, Prince Sultan University (PSU), Riyadh, Saudi Arabia, and also the Leader of the AIDA Laboratory. She has published over 300 publications in high-ranked journals. Her primary research interests include bioinformatics, data mining, and classification using AI models. She received the Best Student Award from the Faculty of Computing, UTM, in 2012, and the Best Research of the Year Award from PSU, from 2013 to 2016. She is an editor of several reputed journals and on a panel of TPC at international conferences.