Journal of Global Information Management
Volume 32 • Issue 1

How Manufacturing Companies Can Improve Their Competitiveness:
Research on Service Transformation and Product Innovation Based on Computer Vision

Yongling Zhang, School of Information and Communication Engineering, North University of China, China
Huaqing Du, International Business School, Hebei International Studies University, China
Tianyu Piao, School of Economics, Changchun University, China
Hongyu Shi, Business School, Henan Normal University, China*
Sang-Bing Tsai, International Engineering and Technology Institute, Hong Kong
https://orcid.org/0000-0001-6988-5829

ABSTRACT

As the global market continues to evolve and competition escalates, the business environment becomes increasingly competitive. How manufacturing companies improve their competitiveness has always been a topic of great concern. Service transformation and product innovation are key factors and are considered to be important ways for enterprises to stand out in the market. Traditional service transformation and product innovation processes often face complex problems, including the diversity of customer needs and fierce market competition. This makes it difficult for companies to accurately capture market opportunities, provide personalized solutions, and respond quickly to changes. At the same time, many companies also face problems with product quality control and production efficiency, which further weakens their competitiveness. It is against this background that the importance of computer vision technology has become increasingly prominent.

KEYWORDS

CGAN, computer vision, EfficientNet, product innovation, service transformation, YOLOv5

INTRODUCTION

As an important topic in the business field, research on service transformation and product innovation has always attracted much attention. Amidst intensifying competition in the global market, companies constantly seek new ways to enhance their competitiveness and meet changing customer needs (Klinker et al., 2020). In this context, service transformation and product innovation have become the two

DOI: 10.4018/JGIM.336485 *Corresponding Author

core strategies driving the development of enterprises. However, despite their evident value, these strategies encounter a series of complex challenges in practical implementation.

On the one hand, service transformation involves the transformation of traditional product-oriented business models into more flexible and personalized service models. This change requires a deep understanding of customer needs, the provision of customized solutions, and their conversion into viable business models (Alt et al., 2018). For example, in the automotive industry, manufacturers are shifting toward service-oriented models like car-sharing services, requiring an overhaul of their business strategies and customer interaction approaches. However, this process involves intricate information processing and decision-making, demanding effective tools for support.

On the other hand, product innovation requires companies to continuously introduce new products to meet the diverse needs of the market. This requires companies to identify market trends in a timely manner and quickly design and launch innovative products (Tian et al., 2024). In the tech industry, for instance, companies like Apple and Samsung regularly innovate their product lines to maintain a competitive lead in the market. Yet, a significant challenge during product development lies in effectively assessing the feasibility and potential market reception of these new products.

At this stage, computer vision technology has emerged as a focal point in research on service transformation and product innovation, providing businesses with powerful tools to address the aforementioned challenges. Through computer vision, companies can better understand their customer needs (Ng et al., 2021), monitor production processes in real time, and accelerate the development and testing of new products. Despite the theoretical potential of computer vision, effectively using its advantages in practical applications to support service transformation and product innovation remains a topic worthy of in-depth study (Han & Yuan, 2023).

Thus, this article proposes the EfficientNet-YOLOv5-conditional generative adversarial networks (CGAN) model, aiming to explore how to fully harness computer vision technology to address practical problems in service transformation and product innovation. This research aims to delve into the potential of computer vision technology in commercial applications, providing enterprises with more innovative solutions and helping them gain a competitive advantage in the ever-changing market (Gao et al., 2023).

This article outlines a series of experimental results that confirm the effectiveness of the EfficientNet-YOLOv5-CGAN model. To present a comprehensive overview of these results, the authors summarize the key findings of the experiments and describe the main metrics used to measure the effectiveness of the model.

First, the authors use Accuracy, Recall, F1 Score, and Area Under the Curve (AUC) as the main indicators to evaluate the model's performance in both image classification and object detection. These metrics collectively reflect the model's ability to correctly identify images and objects. In particular, AUC is an important indicator, particularly to measure the performance of a classification model when handling imbalanced data sets. In this study's tests, the EfficientNet-YOLOv5-CGAN model demonstrated high Accuracy, high Recall, and high F1 Score, along with an excellent AUC value. These results highlight its significant advantages in processing complex images and real-time object detection.

In comparison with existing solutions, the EfficientNet-YOLOv5-CGAN model outperforms traditional computer vision models across multiple metrics. Especially noteworthy is its efficiency and accuracy in processing large-scale data sets and complex scenes. In addition, this model shows significant advantages in providing customized solutions that can better adapt to the diverse needs of different industries and application scenarios.

Through these experimental results, the authors demonstrate the application potential of the EfficientNet-YOLOv5-CGAN model in the fields of service transformation and product innovation, underscoring its significant advancement over existing technologies. The following section will introduce the characteristics and applications of the EfficientNet-YOLOv5-CGAN model, as well as its potential contributions in the fields of service transformation and product innovation.

This article contributes the following points:

- **Introduction of EfficientNet-YOLOv5-CGAN model:** The article introduces the comprehensive application of EfficientNet, YOLOv5 and CGAN models to address practical challenges in service transformation and product innovation. This comprehensive model not only improves the performance of computer vision but allows companies to better understand customer needs, optimize production processes, and accelerate product innovation. EfficientNet provides efficient image classification, YOLOv5 implements real-time target detection, and CGAN supports the generation of virtual prototypes. The synergy among these three models provides enterprises with more powerful tools for addressing complex challenges in their operations.
- **Implications for Commercial Applications:** This research provides practical examples of computer vision applications in the corporate world, supported by actual cases and experimental verification. The authors demonstrate the potential applications of the EfficientNet-YOLOv5-CGAN model in manufacturing enterprises, including areas like quality control, production line monitoring, and virtual product prototype generation. These application examples not only present novel ideas for enterprises but help improve their competitiveness and responsiveness in the market.
- **Lowering the Technical Threshold:** Thus research extends beyond theoretical exploration, aiming to lowering the technical threshold for the application of computer vision technology. The authors provide user-friendly guidelines and methods to help enterprises more easily adopt the EfficientNet-YOLOv5-CGAN model. This approach allows small- and medium-sized enterprises to make full use of computer vision technology to drive service transformation and product innovation without the need for large technical teams.

Section 2 will delve into recent related research. Section 3 will comprehensively present the methodologies employed, including EfficientNet, YOLOv5, and CGAN. Section 4 will be dedicated to an in-depth examination of the authors' experiments, providing specific details and comparative analyses. Lastly, in Section 5, the authors will draw conclusions and summarize the key findings of this study.

RELATED WORK

Digital Transformation of Traditional Production Companies

In the digital economy of the 21st century, traditional manufacturing companies are under pressure to compete with emerging technology companies (Hermes et al., 2020). To remain competitive and adapt to changing market demands, these companies must consider undergoing digital transformation. In this regard, computer vision technology offers great potential (Wang & Dai, 2022). Computer vision is a technology that enables computers to "see" and "understand" images and video content.

Using deep learning models like convolutional neural networks (CNNs), manufacturers can perform real-time inspection and quality control of products on the production line, reducing production defects and improving productivity (Shahid et al., 2020). For example, an automotive manufacturer uses computer vision technology to capture the component installation process on the production line with a high-resolution camera and analyzes it with algorithms to ensure the correct installation of each component (Riba et al., 2020). In addition to production process optimization, computer vision offers companies a unique way to interact with consumers. For example, some furniture manufacturers use augmented reality (AR) technology, allowing consumers to preview how their furniture will look in a real-life environment through their phone's camera. This not only provides a novel shopping experience but facilitates better consumer understanding of the product.

Despite the tremendous opportunities that computer vision presents for manufacturing companies, its implementation and application still face a number of challenges (Sarker, 2021). First, for effective use of this technology, companies need to conduct a significant amount of initial research and investment. In addition, data management and analysis become challenging due to the large amount of image and video data that needs processing and storage. Finally, traditional manufacturing companies need to invest a lot of time and resources in the digital transformation process.

New Product Development

With the increasing diversity of consumer demands, manufacturing companies must continue to innovate to maintain their leading position in the market. In this process, computer vision technology provides new perspectives and ideas for product development (Tang et al., 2020).

By leveraging computer vision, companies can gain insights into consumer preferences and behaviors that were previously difficult to obtain. For example, by analyzing images uploaded by users on social media, companies can better understand consumer preferences and trends. Using CNN models, these images can be classified, labeled, and analyzed to provide valuable input for product design and marketing.

Additionally, computer vision can play a role in the prototype testing phase of a product. By using computer vision, companies can monitor consumer reactions to new product prototypes in real time and adjust accordingly (Han et al., 2022). For example, when testing a new user interface design, an electronics manufacturer used eye-tracking technology to record consumers' eye movement paths to optimize the design (Arnab et al., 2021).

The application of this approach is not without its challenges. For starters, companies rely on large amounts of image data uploaded by consumers. This not only involves privacy and data protection issues but the need to ensure the quality and authenticity of the data (Graham et al., 2021). Second, the processing and analysis of data require highly specialized knowledge and skills, which can be a barrier for some small- or medium-sized businesses. Finally, even if the results of data analytics provide valuable insights, it can be a challenge to translate these insights into actual product innovations.

A Study on the Application in the Service Industry

With the wide application of digital technology, many traditional processes in the service industry are beginning to be disrupted. In this context, computer vision, with its unique advantages, brings unprecedented opportunities and challenges to the service industry.

In the restaurant industry, computer vision is used in automated ordering and checkout systems (Sarker, 2021). Customers can order food by scanning an image of the food item or using an AR menu without interacting with a server. Additionally, hotels and resorts are beginning to use facial recognition technology to offer customers a seamless check-in experience.

In healthcare, computer vision plays a key supporting role. For example, doctors can use image recognition technology to aid in diagnosis, especially in dermatology or radiology image analysis. In addition, patients are able to perform some basic health checks at home using computer vision-based applications (Fang et al., 2020).

To implement these applications, research can employ recurrent neural network (RNN) models. RNNs (Yu et al., 2019) are particularly well-suited for processing and analyzing sequential data, such as speech or text, making them ideal for use in service scenarios, especially when addressing customer feedback or queries.

The use of computer vision in the service industry brings many advantages; however, there are still a number of problems associated with its application. In practice, misrecognition or incorrect feedback may occur, which may lead to customer dissatisfaction. In addition, consumers accustomed to traditional service processes may require a certain period of adaptation to new computer vision-based services.

Supply Chain Management

Supply chain management is the core of any manufacturing and distribution organization. Effective supply chain management not only improves efficiency but also reduces costs (Ulhaq et al., 2020). The application of computer vision technology has revolutionized supply chain management, offering transformative solutions for enhanced efficiency and cost reduction.

In warehouse management, computer vision is used to automatically identify and track goods (Chang & Chen, 2020). Through the use of smart cameras and sensors, warehouse management systems can obtain real-time information on the location and status of goods, optimizing the storage and retrieval process. In addition, this technology can be used for automated cargo inspection, ensuring the quality and integrity of goods. In logistics and distribution, computer vision aids in real-time tracking of vehicles and goods (Juma et al., 2019). Using the You Only Look Once (YOLO) model, the system can detect and track vehicles in real time, ensuring timely and secure delivery of goods to their destinations.

Although the adoption of computer vision in supply chain management has delivered significant benefits, challenges and limitations persist in practical applications. For complex supply chain systems, such as multinational supply chains, the implementation and maintenance of computer vision technologies can be complex and expensive. In addition, ensuring the accuracy and reliability of the system is a major challenge, especially in harsh environments or unstable networks.

The authors emphasize the potential of computer vision technology in the digital transformation of traditional manufacturing enterprises, new product development, application research in service industries, and supply chain management. The selection of computer vision in these fields is not unreasonable given its unique advantages. Compared with other technologies, computer vision provides a more intuitive and efficient approach to analyzing and processing large amounts of visual data.

In terms of digital transformation, computer vision is instrumental in automating product identification and classification, improving production efficiency and quality control. These capabilities are difficult to achieve with alternative technologies, especially in scenarios where large-scale visual information needs to be processed. In the domain of new product development, computer vision facilitates the rapid identification of market trends and consumer preferences by analyzing images of existing products, offering an intuitive visual analysis method for flexible responses to market shifts and innovative product design. In the service industry, computer vision provides an enriched service experience by providing personalized offerings through visual analysis of customer behavior. For example, in the retail industry, monitoring customers' shopping behavior and preferences enables the delivery of more tailored shopping recommendations. Finally, in supply chain management, computer vision technology can monitor and optimize logistics processes in real time. It can identify and track products, ensuring efficiency and transparency in the supply chain—an act that challenges traditional technologies.

Thus, the unique advantages of computer vision technology make it particularly adept at dealing with challenges in these areas, underscoring the main reason behind the authors' research focus.

METHOD

Network Overview

The EfficientNet-YOLOv5-CGAN model studied is a comprehensive architecture that integrates image classification, real-time object detection, and conditional image generation capabilities. First, EfficientNet is used for efficient image classification and feature extraction, ensuring swift and accurate analysis of large amounts of data. Next, the introduction of YOLOv5 equips the model with fast object detection capabilities in real-time environments, particularly important for applications like real-time defect detection on production lines. CGAN, a variant of generative adversarial networks, is adept at creating realistic virtual prototypes based on conditional inputs. In this model, CGAN takes the outputs

from EfficientNet and YOLOv5 as conditional inputs to generate virtual prototypes. This integration is key to the authors' approach, allowing for the generation of highly relevant and customized virtual prototypes. The prototypes generated by CGAN reflect both the image classifications provided by EfficientNet and the object details identified by YOLOv5, ensuring close alignment with customer preferences and real-world objects. CGAN adds the ability to generate content according to specific conditions, providing technical support for various personalized and customized needs. This versatile model structure ensures efficiency and accuracy, crucial for practical applications in diverse scenarios.

To provide context, the current research was conducted in the Pytorch 1.12.1 environment. All model training and evaluations were preformed on NVIDIA RTXA4000 GPUs. In addition, transfer learning plays a key role because EfficientNet-YOLOv5 shares most weights with YOLOv5. This means that we can benefit from the performance of YOLOv5, leading to a significant reduction in training time. This detail not only conserves computing resources but also expedites the deployment of models in practical applications.

Figure 1 presents the EfficientNet-YOLOv5 framework and Figure 2 showcases its architecture. The design of this framework is inspired by a comprehensive analysis of a dataset with imbalanced categories and fine-grained data, common challenges in practical applications. To enhance its effectiveness in handling these challenging situations, the authors made improvements to the YOLOv5 model, especially to make it more effective in handling fine data and imbalanced categories. The decision to replace the YOLOv5 backbone network with EfficientNet was driven by EfficientNet's unique ability to improve model accuracy while maintaining high efficiency, an important advantage when processing extensive image data.

The exploration of the complex YOLOv5 series involved adding or removing components from the network and observing their functions, a process that played a pivotal role in the experiments. During the experimental stage, the authors split the provided data set into an 8:2 ratio and improved the model performance from 36% to 40.53% through data enhancement and parameter adjustment. Subsequently, offline data augmentation methods were used to generate new datasets for extensive experiments on both YOLOv5 and EfficientNet-YOLOv5. Ultimately, the results indicate that the EfficientNet-YOLOv5 model achieves higher performance levels on the new dataset. Furthermore, the authors found that leveraging the strengths of different models and integrating them together has demonstrated positive prospects for achieving improved scores.

Data augmentation aims to expand the dataset, enabling the model to acquire a more comprehensive understanding about marine microalgae images in different environments. This comprehensive learning helps improve the robustness of the model, allowing it to perform well in a variety of complex situations. This study uses both offline data augmentation and a test-time augmentation method to increase the stability and generalization capability of the model.

Figure 1. EfficientNet-YOLOv5 framework

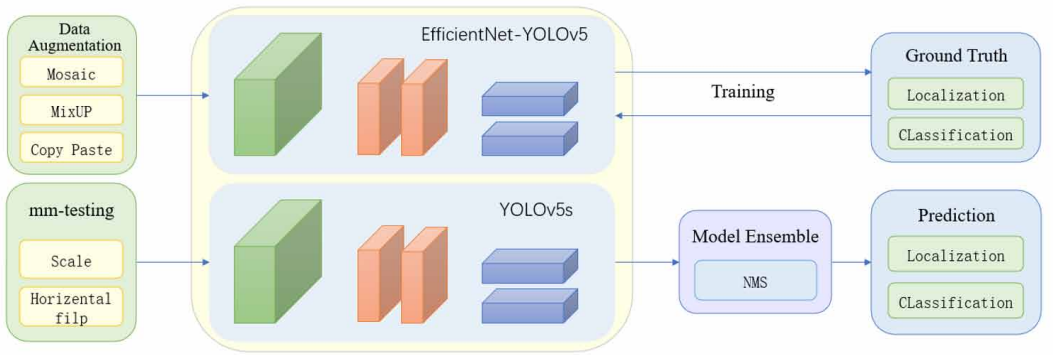
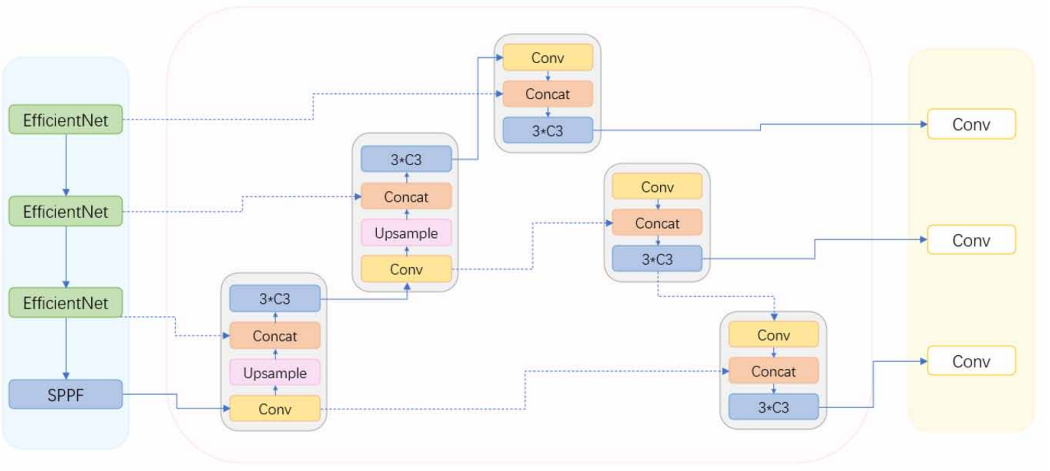


Figure 2. EficientNet-YOLOv5 architecture



Offline data enhancement methods include the addition of random pixels, introduction of Gaussian noise, Cutout random rectangular occlusion, Gaussian blur, and motion blur. The comprehensive application of these methods allows the model to better adapt to a wide array of data situations.

In the experiments, thea authors introduced a test-time enhancement method aimed at improving the stability and generalization ability of the model. The approach includes multiple techniques, such as data augmentation, model ensemble, and dynamic learning rate adjustment.

Data augmentation involves introducing slightly changed input data, enhancing the model's adaptability to new situations. Model ensemble combines the prediction results of multiple models, improving the accuracy and robustness of the overall prediction. Dynamic learning rate adjustment facilitates more efficient learning and during training. These test-time enhancement techniques had a significant impact on model performance. Comparative experiments revealed that models applying the test-time enhancement method had significant improvements in key indicators like Accuracy, Recall, and F1 Score. Especially when dealing with complex and varied data sets, this method effectively improves the model's generalization ability, making it more stable and reliable in practical applications.

EfficientNet

EfficientNet, introduced by Google in 2019, is an innovative CNN structure that stands out due to its unprecedented model efficiency and accuracy. In contracst to traditional neural network structures, it skillfully balances the scaling of network depth, width, and input resolution (Alhichri et al., 2021). This approach allows EfficientNet to provide a much higher level of accuracy for the same computational cost.

The architecture of EfficientNet is designed to strike a balance between computational and parametric efficiency to improve performance (Wang et al., 2020). The idea of this architecture is to continuously expand and deepen the model, using techniques like automated machine learning (AutoML) and reinforcement learning. These techniques play a vital role in searching for and designing the best network architecture.

EfficientNet introduces a compound scaling method to optimize different dimensions of the network, including depth (number of layers in the network), width (number of channels per layer) and resolution (size of the input image). This combined tuning method improves performance without adding excessive computational resources (Huang et al., 2021). Compared to previous network architectures, EfficientNet focuses more on achieving a balance between width and depth to achieve

superior performance. This means that each layer of the network is carefully designed to avoid excessive parameters and computational burden.

In addition, EfficientNet makes extensive use of depth-separable convolution, a lightweight operation in CNNs that reduces computational requirements while maintaining robust feature extraction performance. It is designed to incorporate reinforcement learning techniques to automatically search for suitable network structures through AutoML, ultimately optimizing performance.

An EfficientNet Layer i can be expressed as follows: Y_i is a result of applying the operator $\mathcal{F}_i$ to the input tensor X_i, where Y_i represents the output tensor. The dimensions of the input tensor are given as $\left\langle H_i, \hat{W}_i, C_i \right\rangle$[1], where H_i and W_i correspond to the spatial dimensions and C_i represents the channel dimension. An EfficientNet N can be represented as a composition of multiple layers, denoted as $\mathcal{N}$, where each layer $\mathcal{F}_j$ is combined using the ◉ operator, starting from the initial input tensor X_1. In practical terms, EfficientNet layers are often organized into several stages. All layers within each stage share the same architectural characteristics. For instance, like ResNet, which has five stages, all layers within a single stage in an EfficientNet have the same convolutional type (except the first layer performs down-sampling). Consequently, EfficientNet is defined as:

$$\mathcal{N} = \bigotimes_{i=1\ldots s} \mathcal{F}_i^{L_i} \left(X_{\langle H_i, W_i, C_i \rangle} \right) \tag{1}$$

where $F_i^{L_i}$ denotes layer F_i is repeated L_i times in stage i, and $\left\langle H_i, W_i, C_i \right\rangle$ denotes the shape of input tensor X of layer i.

In contrast to typical EfficientNet designs, which primarily concentrate on optimizing the architecture of individual layers $\left(F_i\right)$, model scaling aims to extend the network's length (L_i), width (C_i), and/or resolution (H_i, W_i) while keeping the baseline network's predefined features $\mathcal{F}_i$ constant. By keeping $\mathcal{F}_i$ fixed, model scaling simplifies the process of designing networks for various resource constraints. However, it presents a substantial challenge in exploring different values for L_i, C_i, H_i, and W_i for each layer. To further narrow the design possibilities, the authors impose a requirement that all layers must be scaled uniformly with a consistent ratio. The objective is to maximize the model's accuracy under specified resource constraints. This objective can be framed as an optimization problem:

$$\begin{aligned} &\max_{d,w,r} Accuracy\left(\mathcal{N}\left(d,w,r\right)\right) \\ &s.t. \mathcal{N}\left(d,w,r\right) = \bigoplus_{i=1\ldots s} \hat{\mathcal{F}}_i^{d\cdot\hat{L}_i} \left(X_{\langle r\cdot\hat{H}_i, r\cdot\hat{W}_i, w\cdot\hat{C}_i \rangle} \right) \end{aligned} \tag{2}$$

where d, w, r are coefficients for scaling network width, depth, and resolution. $\hat{\mathcal{F}}_i, \hat{L}_i, \hat{H}_i, \hat{W}_i, \hat{C}_i$ are predefined parameters in the baseline network.

EfficientNet introduces a compound scaling method to tune on different dimensions of the network including depth (number of layers in the network), width (number of channels per layer) and resolution (size of the input image). This combined tuning helps to improve performance without adding excessive computational resources. Compared to previous network architectures, EfficientNet focuses more on the balance of width and depth to achieve better performance. This means that each layer of the network is carefully designed to avoid excessive parameters and computational burden. In addition, EfficientNet makes extensive use of depth-separable convolution, a lightweight operation

in CNNs that helps reduce computational requirements while maintaining better feature extraction performance. It is designed with the help of reinforcement learning techniques to search for suitable network structures through AutoML to achieve the goal of performance optimization.

EfficientNet has had a significant impact across a variety of industries, particularly in the realm of service transformation where there is a growing demand for faster, more accurate, and automated solutions. Unlike traditional image recognition methods that often require significant computational resources, EfficientNet achieves higher recognition accuracy at a lower computational cost. This enhanced performance makes EfficientNet well-suited for real-time application scenarios, especially those demanding higher video analysis.

EfficientNet introduces several important changes to the field of computer vision. First, it redefines the relationship between model size and speed, enabling high-performance visual recognition even on resource-constrained devices. Its balanced network structure gives EfficientNet a strong generalization capability, enabling it to cope with diverse application scenarios. The combination of efficiency and accuracy presents possibilities for various practical application scenarios, promoting rapid advancements within the industry. As technology progresses, EfficientNet is, therefore, expected to bring about further changes and contribute to various industries.

YOLOv5

YOLOv5, developed in 2020 in collaboration with Glenn Jocher, Yash Sen, Alexey Bochkovskiy and the open-source community, is a significant advancement in target detection algorithms. Compared to its predecessor, YOLOv5 utilizes new techniques and improvements to enhance detection accuracy and speed (Zhang & Yin, 2022). The algorithm is built on a single-stage target detection framework that divides images into small grid cells, predicting the objects present in each cell. It utilizes lightweight network architectures like CSPNet and PP-YOLO, and introduces multi-scale detection techniques to improve the detection of objects of different sizes (Liu et al., 2022).

To improve generalization ability, YOLOv5 employs data enhancement techniques like random scaling and color dithering. Self-supervised learning further improves performance by pre-training on unlabeled data (Zhao & Zhou, 2022). The algorithm focuses on lightweight deployment, making it well-suited for real-world applications on embedded devices and mobile platforms. YOLOv5's strengths position it as a powerful solution for efficient, real-time vision tasks across a range of applications, including industrial settings, unmanned systems, security surveillance, and medical diagnostics.

YOLO is an object detection method that predicts multiple objects in a single forward pass of the neural network. It divides the image into a grid, predicting bounding boxes and class probabilities for each cell. The following equations are core mathematical concepts for YOLOv5.

$$B = \sigma\left(t\right) + c \tag{3}$$

where B is the predicted bounding box, σ is the sigmoid activation function, t is the raw box prediction from the model, and c is the top-left corner coordinates of the grid cell.

Bounding boxes are represented by the center of the box in terms of x and y coordinates, as well as the width and height. The model predicts offsets from anchor boxes rather than the actual width and height.

$$B_{wh} = p_{wh} \odot e^{t} \tag{4}$$

where B_{wh} is the width and height of the bounding box, p_{wh} is the dimensions of the anchor box, t is the raw width and height predictions from the model, and $\odot$ is element-wise multiplication.

The objectness score, another component of YOLO, provides a scalar value, indicating the likelihood of an object being present in the predicted bounding box.

$$O = \sigma\left(o\right) \tag{5}$$

where O is the objectness score, σ is the sigmoid activation function, and o is the raw objectness prediction from the model.

For class predictions, YOLOv5 predicts scores for each potential object category. These scores are then converted into class probabilities using a softmax function.

$$P\left(C_k\right) = \frac{e^{s_k}}{\sum_{i=1}^{N} e^{s_i}} \tag{6}$$

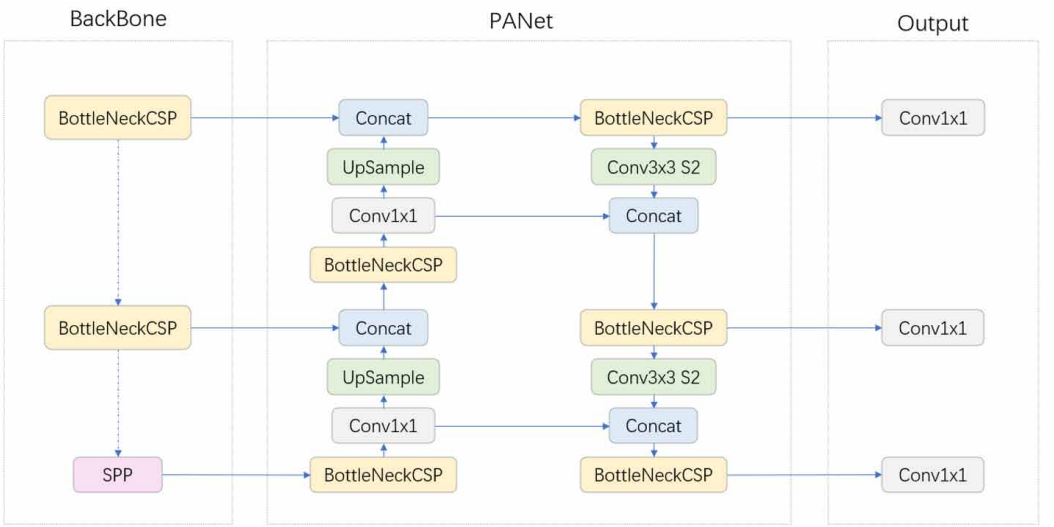
where $P\left(C_k\right)$ is the probability of class k, s_k is the raw score for class s_k from the model, N is the total number of classes, and s_i is the raw score for class i.

YOLOv5's predictions are based on these equations. The bounding boxes are offset and scaled based on grid cells and anchor boxes, ensuring accurate box predictions. The objectness and class scores are normalized using activation functions to give meaningful outputs. The structure of YOLOv5 is shown in Figure 3.

The application of YOLOv5 in manufacturing enterprises has far-reaching effects and significance. Manufacturing companies need to take swift, accurate measures to monitor and identify products on the production line, as well as detect abnormalities in the production process. Fulfillment of this need is critical to maintaining the competitiveness of the organization (Pham et al., 2023). YOLOv5 brings significant benefits to manufacturing companies with its superior fast target detection capabilities.

The application of YOLOv5 in manufacturing enables the automation of quality control on the production line, monitoring the production process in real time. This act ensures product

Figure 3. YOLOv5 architecture



compliance and quality, contributing to a reduction in production costs, improved productivity, and the minimization of manual errors. Beyond operational efficiency, YOLOv5 also ensures that products meet stringent quality standards that satisfy customers' needs.

The application of YOLOv5 is particularly important in this experiment, focused on improving the competitiveness of manufacturing companies. It plays a key role in several critical areas that contribute to operational excellence and improve market position.

First, YOLOv5 can be used to detect product defects on production lines, helping companies to identify and correct potential quality problems in a timely manner. Second, it can monitor logistics in the warehousing process. This improves the efficiency of logistics and inventory management, thereby reducing the waste of resources. Furthermore, YOLOv5's ability to track the transportation process of products ensures their safe and timely delivery to their destinations, aligning with customer needs.

By using YOLOv5's target inspection technology, companies can improve overall productivity, reduce operational costs, and deliver high-quality products. These factors will help companies to stand out in a competitive market, transforming their services and innovating their products. Thus, they can increase their competitiveness and market share.

CGAN

CGAN represent an extension of the original generative adversarial networks (GAN). The GAN consists of a generator (Loey et al., 2020) responsible for creating data from a random noise and a discriminator tasked with distinguishing between real data and generated data. In this adversarial process, the generator constantly tries to generate more realistic data, while the discriminator attempts to distinguish between the generated data and real data.

The core idea of CGAN is to introduce additional conditional information to both the generator and the discriminator. These conditions, often denoted as "y," can take the form of labels, text descriptions, or other relevant information. Unlike traditional GANs where the generator uses only random noise "z," CGAN's generator takes into account both the random noise "z" and the condition "y" to generate the desired data (Torkzadehmahani et al., 2019). Similarly, the discriminator will consider both the data and condition "y" to perform the discrimination.

To gain a deeper understanding into the mechanics of conditionalization, this study will explore into the mathematical model of CGAN, gaining a more intricate understanding of how it works at the formula level (figure 4).

CGAN, an extension of the traditional GAN framework, uses additional conditional information to allow us to generate data with specific attributes. Equation 7 is the traditional GAN framework:

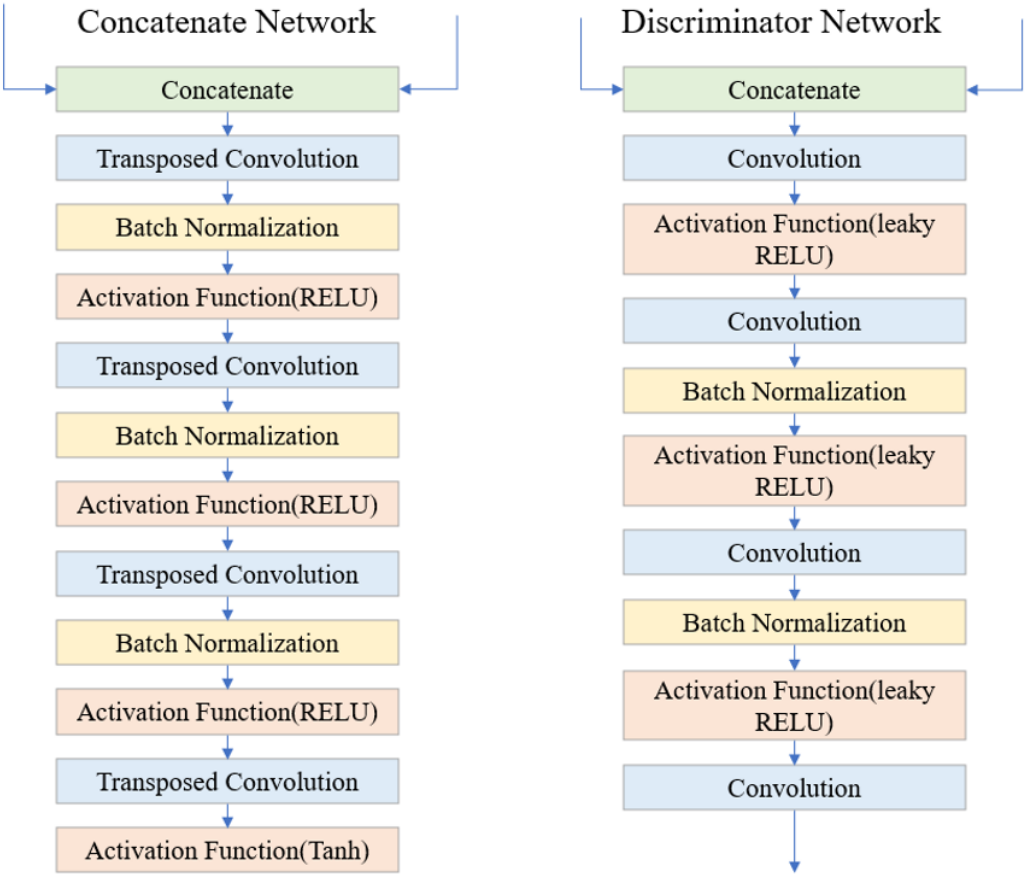
$$\min_{G}\max_{D} V\left(D,G\right)=\mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\log D\left(x\right)\right]+\mathbb{E}_{z\sim p_{z}(z)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right] \tag{7}$$

where G is the generator network, D is the discriminator network, x is data from the real distribution, z is noise input to the generator, and p_{data} and p_{z} are the data and noise distributions, respectively.

Equation 7 represents the value function in the vanilla GAN setting. The generator G tries to minimize this value while the discriminator D tries to maximize it. For CGANs, we condition both the generator and discriminator on some additional information y. Therefore, the generator and discriminator become functions of both the noise z and the condition y.

$$G:\left\{z,y\right\}\rightarrow x$$

Figure 4. Structure of the proposed CGAN network



$$D:\{x,y\}\rightarrow[0,1] \tag{8}$$

where z is noise, y is the conditional information, and x is the generated data. x is data (either real or generated), y is the conditional information, and the output is the probability that x is real. Given this setting, the objective of a CGAN is adapted as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D\left(x\#\ y\right)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1-D\left(G\left(z\middle|y\right)\right)\right)\right] \tag{9}$$

where y is the conditional information. The other symbols maintain the meanings from the previous equations.

Within a CGAN, the objective aligns with traditional GANs, but with the distinction that both data generation and discrimination processes are conditioned on additional information, denoted as y. The training process in CGAN alternates between two phases: updating the discriminator while keeping the generator fixed and vice versa. This approach ensures that both the generator and

discriminator improve in tandem, leading to the generation of data that becomes increasingly difficult to distinguish from real data, given the condition y.

In the context of modern production enterprises, improving competitiveness hinges on meeting consumers' individualized needs and fostering rapid innovation. At this point, the intervention of CGAN brings great potential value to production enterprises (Gonzalez et al., 2022). CGAN can generate corresponding image content based on specific conditions or labels, serving a powerful tool for product design and innovation. In turn, CGAN provides a powerful computer vision tool for production enterprises to not only realize the needs of customers but also stay ahead of in terms of innovation of their products, thereby gaining greater competitiveness in the market.

Using CGAN in production companies introduces a transformative capability, enabling the quick generation of sketches for numerous product designs. This tool creates a rich and creative design space, allowing for product development teams to iterate and optimize designs swiftly. Moreover, this approach significantly reduces the time from concept to prototype, accelerating the process of getting products to market (Naderi et al., 2022).

Furthermore, it CGAN offers great possibilities for customized production. Traditionally, customized production is often accompanied by high costs and time investments. However, with CGAN, companies can generate corresponding product designs based on consumers' descriptions or preferences. This facilitates true on-demand production, satisfying consumers' pursuit of personalization.

RESULTS

Datasets

ImageNet Dataset (Image-Net, 2023)

ImageNet is an expansive image dataset widely used in computer vision and deep learning research. Containing millions of high-resolution images covering thousands of categories, each category is represented with hundreds to thousands of images. This dataset has become a standard in the field, serving as a benchmark for evaluating the performance of various image classification algorithms.

Best known for image categorization, ImageNet involves assigning each image to its corresponding object or scene category. Thus, it has become the de facto benchmark for image classification. Each image in the dataset is manually labeled with one or more category labels, forming the basis for training and evaluating image classification algorithms.

In addition to image classification, the ImageNet dataset is widely used for training and evaluating a variety of deep learning algorithms. It finds applications in target detection, image generation, and migration learning, contributing to advancements in diverse aspects of computer vision research. By providing a rich and diverse image resource, the ImageNet dataset not only promotes research but also propels technological advances in the field of computer vision.

Common Objects in Context (COCO) Dataset (Cocodataset, 2019)

The COCO dataset is widely used in computer vision and deep learning research, designed to support the advancement of computer vision tasks, including target detection, object segmentation, and scene understanding. Real-time target detection algorithms like YOLOv5 are often trained and tested using the COCO dataset to validate their performance. With more than a million images, each is manually labeled with detailed annotations like the target's bounding box for target detection, pixel-level object segmentation masks for semantic segmentation, and the object's category labels.

The COCO dataset has emerged as an important benchmark in the field of computer vision. It is vital to improving the performance and fostering innovation in image analysis tasks. Researchers and engineers can utilize the COCO dataset to train and evaluate various types of computer vision

algorithms, ranging from target detection to image segmentation. This dataset improves the understanding and processing of real-world images.

Amazon Product Review Dataset (Amazon, 2018)

This dataset is built on Amazon product reviews and serves as a comprehensive resource for researching, analyzing, and developing applications and algorithms related to product reviews. It details consumer reviews, ratings, review texts, and other related information like product ID, user ID, and review timestamps across different Amazon products. The review texts mainly contain user descriptions and opinions about the products.

This dataset is valuable for researchers aiming to develop machine learning models aimed at gaining insights into users' product perceptions, predicting sales trends, or providing product recommendations. Beyond its use in the business domain, these datasets play a crucial role in academic research and data science projects, providing a rich source of information about products and consumer perspectives.

With this information, companies can gain insights into market trends and optimize their products and services to better meet consumer needs. It is imperative, however, that these datasets are used in strict compliance with relevant regulations and data use policies to ensure ethical and responsible use of information.

Kaggle Dataset (Kaggle, 2020)

Kaggle is a quality control dataset that covers a wide range of industries and fields, including manufacturing, medical devices, and electronics. It brings together a large amount of information related to product quality, defects, and inspections, including image data, text data, sensor data, and a variety of other quality control-related data forms. Particularly for image datasets, Kaggle plays a key role in several application scenarios, such as defect detection, quality analysis, and production optimization.

Researchers can comprehensively utilize these datasets to further improve quality control processes by developing predictive models that more accurately identify potential product defects. The datasets contain not only defective images of products but also normal images, which provide a valuable resource for researching and developing image-based defect detection algorithms. The objective is to continually improve product quality and production efficiency.

Kaggle, acting as a collaborative platform, allows data scientists and domain experts from around the world to work together in researching and solving quality control problems. It provides robust support and opportunities for advancements in this critical area, contributing to progress and innovation in quality control processes across the globe.

Experimental Details

In this article, four data sets are selected for training. The training process is as follows:

Step One: Data Processing

Extract sustainability and climate change-related data from ImageNet dataset, COCO dataset, Amazon Product Review dataset, and Kaggle dataset.

Step Two: Model Training

The training process of the EfficientNet-YOLOv5-CGAN module is as follows:

- Obtain the best hyperparameter settings by training the model on the training set and finetuning it on the validation set.

- Adjust the hyperparameters of the model through cross-validation and grid search techniques to improve the performance of the model.
- This experiment adjusts the learning rate of the EfficientNet model, the number of heads in the YOLOv5 model, and the number of layers in the CGAN model.

Step Three: Model Evaluation

Evaluate the trained model using the test set, including calculating metrics like Accuracy, Recall, F1 Score, AUC, Parameters, Inference Time, Flops, and Training Time.

Step Four: Result Analysis

It becomes evident that the EfficientNet-YOLOv5-CGAN model surpasses traditional models after contrasting the performance evaluation measures of the EfficientNet-YOLOv5-CGAN model with conventional models like logistic regression, decision trees, and random forests. An examination of the errors and uncertainties in the BERT-MHA-DNN model across various categories and samples reveals that the model's predictions are notably resilient to factors like data quality and annotation errors. Additionally, scrutinizing the EfficientNet-YOLOv5-CGAN model's ability to generalize to new datasets or real-world scenarios through methods like cross-validation demonstrates a consistent, reliable performance across diverse datasets and scenarios.

Accuracy:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}} \tag{8}$$

where *TP* represents the number of true positives, *TN* represents the number of true negatives, *FP* represents the number of false positives, and *FN* represents the number of false negatives.

Recall:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \tag{9}$$

where *TP* represents the number of true positives and *FN* represents the number of false negatives.

F1 Score:

$$\boldsymbol{F1\_Score} = 2 * \frac{\boldsymbol{precision} * \boldsymbol{recall}}{\boldsymbol{precision} + \boldsymbol{recall}} \tag{10}$$

AUC:

$$\boldsymbol{AUC} = \int_0^1 \boldsymbol{ROC}\left(\boldsymbol{x}\right)\boldsymbol{dx} \tag{11}$$

where *ROC(x)* represents the relationship between the true positive rate and the false positive rate when *x* is the threshold.

Parameters(M):

Count the number of adjustable parameters in the model (in millions).

Inference Time(ms):

Measure the time required for the model to perform inference (in milliseconds).

Flops(G):

Count the number of floating-point operations required for the model to perform inference (in billions).

Training Time(s):

Measure the time required for the model to train (in seconds).

Experimental Results and Analysis

To evaluate the contribution of the different components of the model to its overall performance, the authors conducted an ablation study on four models (i.e., NasNet, FBNet, GhostNet, and EfficientNet) and four datasets (i.e., ImageNet, COCO, APR, and Kaggle). Table 1 and Figure 5 list the experimental results of the study.

First, it is evident that EfficientNet's accuracy outperforms other models across all datasets. Specifically, EfficientNet achieves 93.78% accuracy on the ImageNet dataset, surpassing NasNet (94.76%), FBNet (93.28%) and GhostNet (87.26%). EfficientNet also performs well on the COCO dataset, Amazon Product Review dataset, and Kaggle dataset, achieving accuracies of 95.49%, 90.65%, and 91.89%, respectively, which outperforms other models.

This performance underscores EfficientNet's enhanced generalization ability. Moreover, EfficientNet excels in Recall, F1 Score, and AUC metrics, indicating its superior overall performance. The consistently high values for Recall, F1 Score, and AUC further support its model performance, especially in terms of higher recall rates and overall performance of the model.

EfficientNet performance extends beyond accuracy, consistently demonstrating excellence across various metrics on multiple datasets. This evidence promotes its significant advantage in image categorization tasks. Figure 5 visualizes the table content, highlighting the performance advantage of EfficientNet over other models.

As shown in Table 2, the authors evaluated four image generation models (Pix2Pix, CycleGAN, GAN, and CGAN) on different datasets to understand their performance in image generation tasks. The authors focused on four performance metrics: (1) Accuracy; (2) Recall; (3) F1 Score; and (4) AUC.

First, CGAN performs well in terms of Accuracy. On the ImageNet dataset, CGAN achieves 96.22% accuracy, which is significantly higher than models like Pix2Pix (93.2%), CycleGAN (94.12%), and GAN (93.36%). This means that CGAN has the ability to generate target images with more accuracy, providing higher quality in image generation tasks.

CGAN also excels in terms of Recall. Achieving Recall rates of 92.4% on the ImageNet dataset and 93.01% on the Kaggle dataset, CGAN demonstrates its capability to capture target images, ensuring that the generated images mirror the originals.

At the same time, CGAN achieves significant advantages in terms of F1 Score and AUC. Its high scores in those metrics underscore its superiority in the comprehensive evaluation of model performance, especially in terms of high Recall and overall model performance.

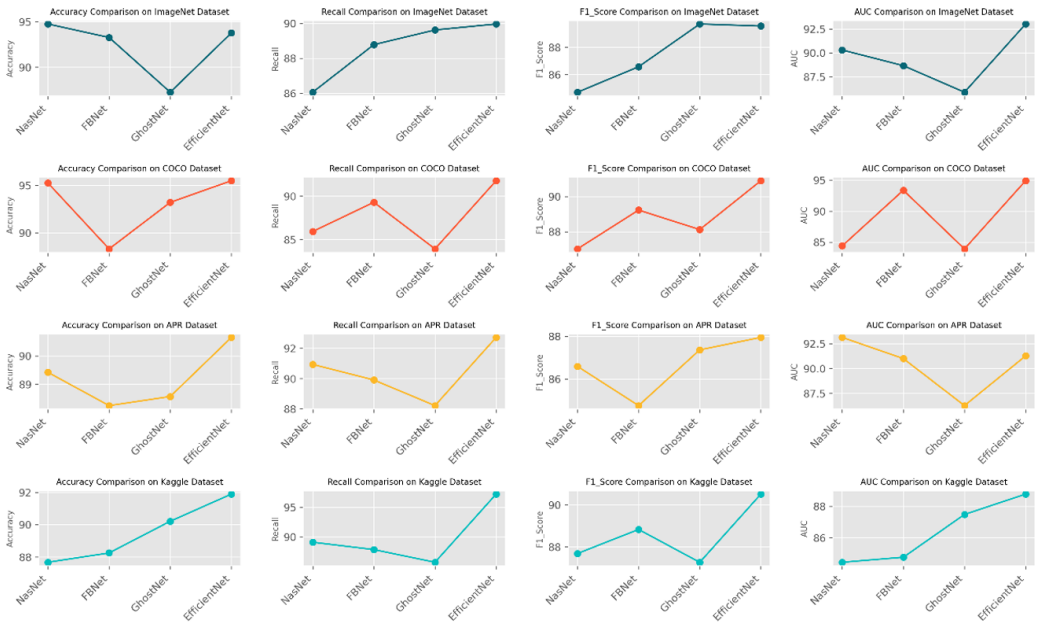
CGAN consistently performs well across multiple datasets, showcasing higher Accuracy, Recall, F1 Score, and AUC. These performance metrics underscore CGAN's significant advantages in image generation tasks. Figure 6 visualizes the table content, further highlighting the performance advantages of CGAN over other models. These results emphasize CGAN as a powerful tool for image generation, offering strong support for further research and diverse applications in the field of image generation.

As shown in Table 3, the authors evaluated the performance of several research teams (Esteva, Chou et al. 2021; García-Morales et al., 2021; Russell, Bvuma et al., 2001; Ulhaq et al. 2020; Yuan et al., 2021) across different datasets. The authors incorporated the work of Lundvall (2015) to benchmark the performance of deep learning models against their own model. They focused on four performance metrics: (1) Accuracy; (2) Recall; (3) F1 Score; and (4) AUC. The metrics provided valuable insights, providing a detailed analysis of performance differences between the models.

Table 1. Ablation experiments on the EfficientNet module using different datasets

Model	Datasets															
	ImageNet Dataset				COCO Dataset				Amazon Product Review Dataset				Kaggle Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
NasNet	94.76	86.06	84.71	90.3	95.25	85.94	87.02	84.41	89.42	90.94	86.59	93.13	87.67	89.1	87.68	84.45
FBNet	93.28	88.78	86.58	88.66	88.31	89.31	89.23	93.35	88.25	89.91	84.76	91	88.25	87.87	88.83	84.77
GhostNet	87.26	89.62	89.72	85.89	93.21	83.93	88.12	83.93	88.57	88.22	87.36	86.27	90.21	85.73	87.26	87.48
EfficientNet	93.78	89.96	89.57	93.01	95.49	91.8	90.9	94.88	90.65	92.72	87.94	91.27	91.89	97.19	90.52	88.77

Figure 5. Comparison of NasNet, FBNet, GhostNet and EfficientNet model performance on different datasets



Regarding Accuracy, the model achieves 95.46% accuracy on the ImageNet dataset. In comparison, other research teams' models have lower accuracy, reaching a maximum of 96.02% (Ulhaq et al., 2020). The model also performs well on the COCO dataset and Kaggle dataset, achieving 94.72% and 97.53% accuracy, respectively, which is much better than other models. This indicates that the authors' model has achieved a significant lead in the image classification task.

In terms of Recall, the model also excels. On the ImageNet dataset, the authors' model achieves a recall of 93.75%, outperforming competing research teams' models. Similarly, on the COCO dataset and Kaggle dataset, their model achieves a recall of 94.86% and 94.35%, respectively, which is also a significant advantage over other models. This highlights the model's ability to of capture the features of the target image.

In terms of F1 Score, on the ImageNet dataset, the model obtains an F1 Score of 92.24, significantly surpassing other models. On the COCO dataset and Kaggle dataset, the model achieves F1 Scores of 93.48 and 92.16, respectively, outperforming competing models. This indicates the model's advantage in balancing Precision and Recall.

The AUC reinforces the outstanding performance of the model. Across all datasets, the model achieves the highest AUC values: 93.16 (ImageNet), 95.17 (COCO), and 91.89 (Kaggle). This again confirms the excellent performance of the model in image classification tasks.

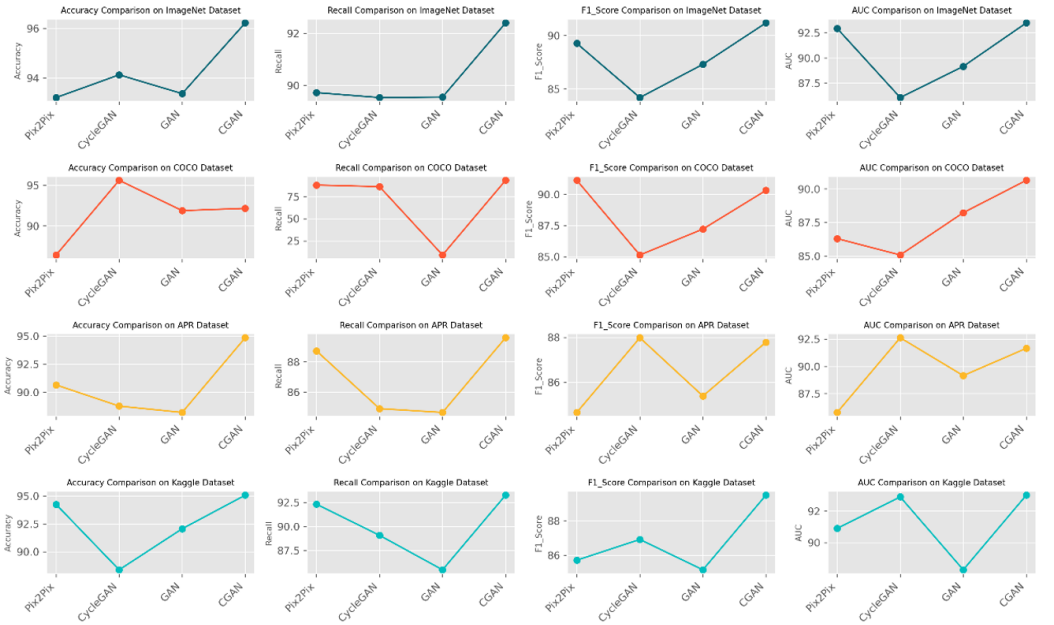
By comparing the performance of different models on various performance metrics, it is clear that the authors' model outperforms other models in terms of Accuracy, Recall, F1 Score, and AUC. This positions the model as a superior solution for the image classification task, surpassing the performance of other models. Not only do these results emphasize the performance advantages of the model, but they provide strong support for further research and applications in the field.

Additionally, this study provides detailed data on model parameter counts (Parameters), computational complexity (Flops), inference time (Inference Time), and training time for different research methods across various datasets. These data-driven comparisons reveal the computational and performance characteristics of different models, revealing their applicability and efficiency across different datasets.

Table 2. Ablation experiments on the CGAN module using different datasets

Model	Datasets															
	ImageNet Dataset				COCO Dataset				Amazon Product Review Dataset				Kaggle Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Pix2Pix	93.2	89.72	89.25	92.92	86.4	87.8	91.12	86.3	90.65	88.73	84.65	85.76	94.24	92.31	85.71	90.9
CycleGAN	94.12	89.53	84.2	86.05	95.62	86.01	85.13	85.08	88.77	84.9	87.98	92.62	88.38	89.07	86.91	92.9
GAN	93.36	89.55	87.3	89.15	91.88	9.039	87.21	88.23	88.21	84.65	85.38	89.14	92.06	85.48	85.15	88.27
CGAN	96.22	92.4	91.16	93.47	92.16	93.13	90.31	90.64	94.84	89.6	87.78	91.65	95.07	93.25	89.48	93.01

Figure 6. Comparison of Pix2Pix, CycleGAN, GAN, and CGAN model performance on different datasets



Analyzing the data reveals notable variations in the number of model parameters across different research methods and datasets. For example, the model of Esteva, Chou et al. (2021) has 530.88M parameters on the ImageNet dataset but reduces to 522.77M parameters on the COCO dataset. In contrast, the authors' model maintains close to 340M parameters on both datasets. This suggests that their model has higher parameter efficiency with relatively fewer parameters.

In terms of computational complexity (Flops), Yuan et al. (2021), for example, requires 8.19G Flops on the ImageNet dataset and 7.5G Flops on the COCO dataset. In comparison, the authors' model exhibits lower computational complexity, hovering around 3.5G Flops on both datasets. This indicates that their model has superior efficiency in terms of computational complexity.

Regarding inference time, variations are shown among different data models across datasets. For instance, Ulhaq et al. (2020) requires 9.79ms inference and 8.52ms inference times on the Amazon Product Review dataset and the Kaggle dataset, respectively. In contrast, the authors' model achieves an inference time of approximately 5.6ms on both datasets, which is significantly faster than the other models. This indicates that their model is more competitive in terms of inference speed.

Training time, an important indicator of efficiency, reveals additional distinctions among models. García-Morales et al. (2021) requires 13.42 seconds of training time on Kaggle dataset, while Ulhaq et al. (2020) requires only 8.52 seconds on Kaggle dataset. In comparison, the authors' model takes around 5.6 seconds on both datasets, showing higher training efficiency.

By comparing the number of parameters, computational complexity, inference time, and training time, the authors' model excels. Demonstrating superior parameter efficiency, computational efficiency, inference speed, and training efficiency, the model showcases outstanding performance across dimensions. These findings provide a solid foundation for the widespread use of the model in real-world applications, spanning computer vision, natural language processing, and other AI domains.

The performance advantages will have a positive impact on future deep learning research and engineering applications. They will also drive the continuous advancement of AI technology. Figure 7 & 8 as well as Table 4 represent the data, providing a clearer appreciation of the performance disparities among the models.

Table 3. Comparison of different models in different indicators comes from the ImageNet dataset, COCO dataset, Amazon Product Review dataset, and Kaggle dataset

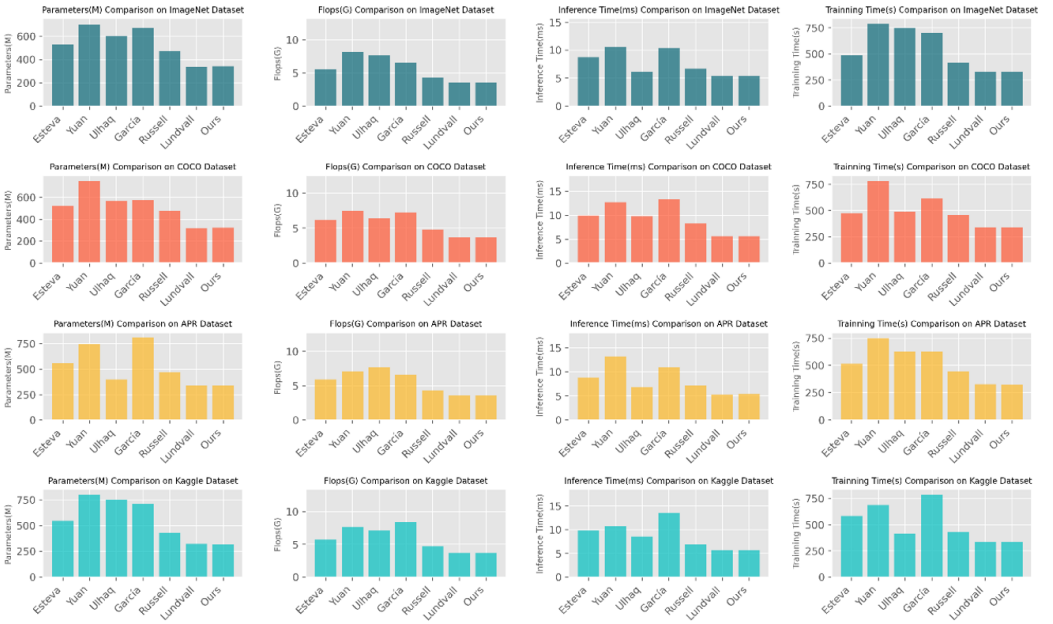
Model	Datasets															
	ImageNet Dataset				COCO Dataset				Amazon Product Review Dataset				Kaggle Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Esteva et al. (Esteva et al., 2021)	91.22	92.28	87.15	92.37	85.89	92.96	86.42	84.28	85.59	89.47	88.91	85.44	93.79	91.5	87.73	91.82
Yuan et al. (Yuan et al., 2021)	92.05	88.35	89.45	83.96	88.29	91.52	88.06	89.21	94.52	84.93	90.16	93.25	95.7	91.5	86.4	91.84
Ulhaq et al. (Ulhaq et al., 2020)	96.02	84.5	83.79	88.37	89.75	92.63	89.48	93.4	93.79	92.44	86.92	84.86	87	87.26	89.97	84.97
García et al. (García et al., 2021)	87.59	90.59	87.48	92.69	87.71	84.2	84.79	90.02	91.24	84.27	91.15	92.72	86.53	91.93	90.44	85.96
Russell et al. (Russell et al., 2001)	93.16	91.8	90.91	85.62	89.32	93.48	89.14	83.9	90.63	88.14	86.14	87.28	94.96	91.17	87.05	93.02
Lundvall(Lundvall, 2015)	93.08	91.97	84.45	87.23	93.64	88.36	91.15	85.25	95.98	89.65	85.01	92.34	87.74	86.52	88.76	86.33
Ours	95.46	93.75	92.24	93.16	94.72	94.86	93.48	95.17	96.48	94.15	91.93	94.16	97.53	94.35	92.16	91.89

Table 4. Comparison of model performance of different models on different metrics from ImageNet dataset, COCO dataset, Amazon Product Review dataset, and Kaggle dataset

Model	Datasets															
	ImageNet Dataset				COCO Dataset				Amazon Product Review Dataset				Kaggle Dataset			
	Parameters (M)	Flops (G)	Inference Time(ms)	Trainning Time(s)	Parameters (M)	Flops (G)	Inference Time(ms)	Trainning Time(s)	Parameters (M)	Flops (G)	Inference Time(ms)	Trainning Time(s)	Parameters (M)	Flops (G)	Inference Time(ms)	Trainning Time(s)
Esteva et al.	530.88	5.58	8.71	487.44	522.77	6.14	9.94	477.09	558.38	5.91	8.8	516.59	544.33	5.69	9.77	584.43
Yuan et al.	699.31	8.19	10.6	787.66	746.59	7.5	12.73	782.04	743.71	7.01	13.11	750.58	797.78	7.66	10.72	686.79
Ulhaq et al.	600.95	7.63	6.16	745.98	568.74	6.39	9.79	489.62	400.2	7.72	6.82	625.26	751.7	7.07	8.52	414.17
García et al.	668.65	6.53	10.35	700.4	573.81	7.23	13.26	613.3	809.56	6.55	10.85	628.53	712.34	8.33	13.42	785.78
Russell et al.	473.28	4.33	6.63	417.76	473.95	4.8	8.25	457.1	468.77	4.33	7.21	447.08	429.13	4.64	6.89	428.7
Lundvall	336.85	3.53	5.33	325.61	317.02	3.65	5.63	338.84	339.44	3.55	5.33	326.34	320.01	3.65	5.6	337.82
Ours	338.93	3.53	5.35	325.19	320.01	3.65	5.6	338.74	339.92	3.54	5.35	325.84	318.5	3.63	5.62	337.11

Figure 7. Visualizations of the results of comparing different models on different metrics come from the ImageNet dataset, the COCO dataset, the Amazon Product Review dataset, and the Kaggle dataset

Figure 8. Visualizations comparing the model performance of different models on different metrics are shown from the ImageNet dataset, the COCO dataset, the Amazon Product Review dataset, and the Kaggle dataset



CONCLUSION AND DISCUSSION

This study focuses on researching service transformation and product innovation based on computer vision. The aim is to enhance the competitiveness of manufacturing enterprises by using advanced computer vision technology and models. The combined use of these models promotes the exploration into the potential of manufacturing companies in terms of service provision and product innovation. Through experiments, the effectiveness of these models in improving product quality, optimizing production processes, and providing innovative services is verified. The research provides manufacturing companies with powerful tools to achieve greater success in a highly competitive market.

The findings of this study offer practical guidelines for manufacturing companies looking to implement advanced computer vision technologies. Companies can integrate the EfficientNet-YOLOv5-CGAN model into their existing systems to enhance quality control, streamline production processes, and foster product innovation. For instance, the model can be applied in intelligent manufacturing settings for real-time defect detection and in supply chain management to improve logistics efficiency.

The fusion model performed well across various data sets in this study. However, certain limitations were identified during the research. Notably, the model's complexity and computational resource requirements may pose challenges to some small or resource-constrained businesses. Additionally, the extensive data and time needed to train and tune models can slow down the implementation of computer vision solutions in some companies. Future research should prioritize addressing these to achieve better model performance and wider applicability. By overcoming these challenges, researchers aim to deliver more reliable and efficient solutions for scientific research and practical applications in the future.

Future research should extend to several key areas to enhance the practical applicability and efficiency of this model. The optimization of the model's architecture to reduce computational overhead and enhance adaptability to diverse datasets will be a primary focus. This will involve refining the model's algorithmic efficiency and exploring lightweight versions suitable for businesses with limited computational resources.

Moreover, future research will explore new application scenarios, particularly in emerging fields like autonomous robotics in manufacturing and predictive maintenance using computer vision. These areas offer substantial potential for leveraging the model's capabilities in dynamic and real-time environments. Additionally, the rapidly evolving landscape of computer vision, marked by advancements in areas like neuromorphic computing and edge artificial intelligence, presents exciting opportunities for future research. These emerging trends could significantly enhance the speed and efficiency of computer vision models, including the authors', presenting novel application scenarios that were previously unfeasible. The authors' goal is to continuously evolve the model to harness these advancements, offering more robust, reliable, and efficient solutions for a wider range of scientific research and practical applications.

ACKNOWLEDGEMENTS

This work was supported by Shanxi Scholarship Council of China (No. HGKY2019078).

REFERENCES

Alhichri, H., Alswayed, A. S., Bazi, Y., Ammour, N., & Alajlan, N. A. (2021). Classification of remote sensing images using EfficientNet-B3 CNN model with attention. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 14078–14094. doi:10.1109/ACCESS.2021.3051085

Alt, R., Beck, R., & Smits, M. T. (2018). FinTech and the transformation of the financial industry. *Electronic Markets*, *28*(3), 235–243. doi:10.1007/s12525-018-0310-9

Amazon. (2018). *Amazon customer reviews dataset.* Amazon. https://registry.opendata.aws/amazon-reviews/

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., & Schmid, C. (2021). Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6836–6846). IEEE.

Chang, S. E., & Chen, Y. (2020). When blockchain meets supply chain: A systematic literature review on current development and potential applications. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 62478–62494. doi:10.1109/ACCESS.2020.2983601

COCO. Common Objects in Context. (2019). [Dataset]. COCO. https://cocodataset.org/

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, *4*(1), 5. doi:10.1038/s41746-020-00376-2 PMID:33420381

Fang, H. S., Wang, C., Gou, M., & Lu, C. (2020). Graspnet-1billion: A large-scale benchmark for general object gras**. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11444–11453). IEEE. doi:10.1109/CVPR42600.2020.01146

Gao, X., Zhou, Z., Wang, L., Shao, G., & Zhang, Q. (2023). Feature selection and text clustering algorithm based on binary mayfly optimization. *Journal of Jilin University Science Edition*, *61*(3), 631–640.

García-Morales, V. J., Garrido-Moreno, A., & Martín-Rojas, R. (2021). The transformation of higher education after the COVID disruption: Emerging challenges in an online learning scenario. *Frontiers in Psychology*, *12*, 616059. doi:10.3389/fpsyg.2021.616059 PMID:33643144

Gonzalez-Sabbagh, S., Robles-Kelly, A., & Gao, S. (2022). *DGD-cGAN: A Dual generator for image dewatering and restoration.* arviv preprint arviv:2211.10026.

Graham, B., El-Nouby, A., Touvron, H., Stock, P., Joulin, A., Jégou, H., & Douze, M. (2021). Levit: A vision transformer in convnet's clothing for faster inference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 12259–12269). IEEE. doi:10.1109/ICCV48922.2021.01204

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. doi:10.1109/TPAMI.2022.3152247 PMID:35180075

Han, K., & Yuan, S. (2023). Short text semantic similarity measurement algorithm based on hybrid machine learning model. *Journal of Jilin University Science Edition*, *61*(4), 909–914.

Hermes, S., Riasanow, T., Clemons, E. K., Böhm, M., & Krcmar, H. (2020). The digital transformation of the healthcare industry: Exploring the rise of emerging platform ecosystems and their influence on the role of patients. *Business Research*, *13*(3), 1033–1069. doi:10.1007/s40685-020-00125-x

Huang, Y., Pan, G., Li, X., Sun, Z., Koyama, S., & Yang, Y. (2021). Mining potential requirements by calculation of user operations. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(6), 1–14. doi:10.4018/JOEUC.293289

ImageNet Large Scale Visual Recognition Challenge. (2009). *Home.* ImageNet. http://www.image-net.org/

Juma, H., Shaalan, K., & Kamel, I. (2019). A survey on using blockchain in trade supply chain solutions. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 184115–184132. doi:10.1109/ACCESS.2019.2960542

Kaggle. (2020). *Statistical process control (quality control).* Kaggle. https://www.kaggle.com/

Klinker, K., Wiesche, M., & Krcmar, H. (2020). Digital transformation in health care: Augmented reality for hands-free service innovation. *Information Systems Frontiers*, *22*(6), 1419–1431. doi:10.1007/s10796-019-09937-7

Liu, W., Quijano, K., & Crawford, M. M. (2022). YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8085–8094. doi:10.1109/JSTARS.2022.3206399

Loey, M., Manogaran, G., & Khalifa, N. E. M. (2020). A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. *Neural Computing & Applications*, ●●●, 1–13. doi:10.1007/s00521-020-05437-x PMID:33132536

Lundvall, B. A. (1985). Product innovation and user-producer interaction. *The Learning Economy and the Economics of Hope*, *19*, 19–60.

Naderi, M., Karimi, N., Emami, A., Shirani, S., & Samavi, S. (2022). *Dynamic-Pix2Pix: Noise injected cGAN for modeling input and target domain joint distributions with limited training data.* arviv preprint arviv:2211.08570.

Ng, C. Y., Law, K. M., & Ip, A. W. (2021). Assessing public opinions of products through sentiment analysis: Product satisfaction assessment by sentiment analysis. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(4), 125–141. doi:10.4018/JOEUC.20210701.oa6

Pham, T. N., Nguyen, V. H., & Huh, J. H. (2023). Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19. *The Journal of Supercomputing*, *79*(8), 8966–8992. doi:10.1007/s11227-022-04979-2 PMID:36619832

Riba, E., Mishkin, D., Ponsa, D., Rublee, E., & Bradski, G. (2020). Kornia: An open source differentiable computer vision library for pytorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3674–3683). IEEE. doi:10.1109/WACV45572.2020.9093363

Russell, E. W., & Bvuma, D. G. (2001). Alternative service delivery and public service transformation in South Africa. *International Journal of Public Sector Management*, *14*(3), 241–265. doi:10.1108/09513550110390819

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420. doi:10.1007/s42979-021-00815-1 PMID:34426802

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 160. doi:10.1007/s42979-021-00592-x PMID:33778771

Shahid, A., Almogren, A., Javaid, N., Al-Zahrani, F. A., Zuair, M., & Alam, M. (2020). Blockchain-based agri-food supply chain: A complete solution. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 69230–69243. doi:10.1109/ACCESS.2020.2986257

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., & Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. *Frontiers in Plant Science*, *11*, 510. doi:10.3389/fpls.2020.00510 PMID:32508853

Tian, S., Li, L., Li, W., Ran, H., Ning, X., & Tiwari, P. (2024). A survey on few-shot class-incremental learning. *Neural Networks*, *169*, 307–324. doi:10.1016/j.neunet.2023.10.039 PMID:37922714

Torkzadehmahani, R., Kairouz, P., & Paten, B. (2019). Dp-cgan: Differentially private synthetic data and label generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. IEEE. doi:10.1109/CVPRW.2019.00018

Ulhaq, A., Khan, A., Gomes, D., & Paul, M. (2020). *Computer vision for COVID-19 control: A survey.* arv preprint arv:2004.09420.

Wang, J., & Dai, Y. (2022). The agglomeration mechanism of network emerging e-commerce industry based on social science. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–16. doi:10.4018/JOEUC.291561

Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. (2020). Multi-label classification of fundus images with efficientnet. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 212499–212508. doi:10.1109/ACCESS.2020.3040275

Yu, X. (2022). Global multi-source information fusion management and deep learning optimization for tourism: Personalized location-based service. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–21. doi:10.4018/JOEUC.294902

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235–1270. doi:10.1162/neco_a_01199 PMID:31113301

Yuan, L., Chen, D., Chen, Y. L., Codella, N., Dai, X., Gao, J., & Zhang, P. (2021). *Florence: A new foundation model for computer vision.* arviv preprint arviv:2111.11432.

Zhang, M., & Yin, L. (2022). Solar cell surface defect detection based on improved YOLO v5. *IEEE Access : Practical Innovations, Open Solutions*, *10*, 80804–80815. doi:10.1109/ACCESS.2022.3195901

Zhao, Y., & Zhou, Y. (2022). Measurement method and application of a deep learning digital economy scale based on a big data cloud platform. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–17. doi:10.4018/JOEUC.295092