IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 November 2024, accepted 14 December 2024, date of publication 18 December 2024,
date of current version 30 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3519718

RESEARCH ARTICLE

E-Commerce Image Enhancement Method Based on Instance Segmentation and Background Replacement

QIANG GAO[1], HUIPING HU[2], AND WEI LIU[1]
[1]School of Information Engineering, Guangzhou Railway Polytechnic, Guangzhou, Guangdong 511300, China
[2]Department of Primary Education, Shangrao Preschool Education College, Jiangxi 334000, China

Corresponding author: Qiang Gao (gaoq11@qq.com)

This work was supported by Guangzhou Railway Polytechnic Newly Introduced Talents Scientific Research Start-Up Project GTXYR2427.

ABSTRACT In the field of e-commerce, the visual presentation of product images is crucial for attracting consumers, improving conversion rates, and enhancing user experience. However, existing image enhancement methods often struggle to balance high-quality visual effects with computational efficiency, especially when handling complex and diverse datasets. This paper proposes a novel image enhancement method that integrates instance segmentation, dominant color detection, background replacement, realistic shadow generation, and logo addition into a unified framework, optimized for e-commerce product images. The method begins with instance segmentation, effectively separating foreground products from the background to provide clear targets for further processing. Dominant color detection ensures visual consistency by extracting the primary colors of the product images. Background replacement techniques improve the aesthetic appeal by replacing the original background with more suitable or attractive scenes. The addition of realistic shadows enhances the three-dimensional appearance of the product, while logo integration strengthens branding and recognition. Experimental results demonstrate that the proposed method significantly improves recognition accuracy, IoU, and mAP for models such as YOLOv5, SSD, and Faster-RCNN, with YOLOv5 showing improvements of 16.66%, 23.28%, and 24.32%, respectively. With an average processing time of 125 ms per image, the method offers a superior balance between performance and computational efficiency, making it suitable for real-time e-commerce applications. The method also holds promise for other domains, including natural landscape photography, medical imaging, and artwork analysis. Future work will incorporate statistical analyses and extend the dataset to include more diverse product categories, aiming to further validate the generalizability and scalability of the method.

INDEX TERMS E-commerce, image enhancement, instance segmentation, background replacement, image recognition.

I. INTRODUCTION

With the rapid growth of e-commerce, high-quality visual content has become indispensable across multiple applications, including product categorization, recommendation systems, search engines, and targeted advertising. These applications rely on robust machine learning models that require vast quantities of labeled image data to perform effectively. However, the e-commerce industry faces significant challenges in collecting diverse, representative, and accurately labeled images, especially when attempting to cover a wide range of product categories, styles, and presentation formats. Datasets such as DeepFashion and the Amazon Product Dataset exemplify these challenges, revealing issues like inconsistent backgrounds, varying lighting conditions, misaligned product angles, and noisy data. Such

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

visual inconsistencies hinder the training of robust machine learning models, making it essential to develop advanced image enhancement techniques.

Data augmentation has emerged as a key solution to these challenges, providing a systematic approach to generating extensive and varied datasets from limited raw images. By applying transformations such as rotation, flipping, color adjustments, random cropping, and noise addition, data augmentation enables models to generalize effectively and adapt to real-world scenarios. For example, rotation simulates different viewing angles, improving the recognition of objects like furniture or electronic devices in diverse orientations. Mirror flipping creates horizontally or vertically flipped images, enhancing model performance on symmetrical products such as shoes or glasses. Adjusting brightness, contrast, and hue accounts for variations in lighting conditions, ensuring adaptability across different photographic settings. Random cropping focuses on essential product features, such as logos or unique designs, while noise addition mimics real-world imperfections like sensor noise or compression artifacts, making models more resilient.

In the context of e-commerce, data augmentation not only enhances model performance but also addresses critical data-related challenges. These techniques mitigate issues such as class imbalance, sparse samples for niche categories, and labeling noise by generating diverse and representative training samples. For instance, they help models focus on crucial product details, reduce overfitting, and improve robustness against unpredictable visual variations. Moreover, data augmentation plays a pivotal role in enhancing downstream applications. In product categorization tasks, it allows models to recognize items under various conditions, improving classification accuracy. In recommendation systems, augmented images enable more precise matching of products to user preferences, while in targeted advertising, they help optimize the placement and effectiveness of ad content.

Overall, data augmentation contributes significantly to the generalization, robustness, and adaptability of machine learning models in e-commerce, improving performance across tasks and enabling a more seamless user experience. By effectively leveraging these techniques, e-commerce companies can address existing challenges, develop smarter and more scalable solutions, and better serve consumer needs while driving business growth.

II. LITERATURE REVIEW

A. DATA AUGMENTATION OVERVIEW

Data augmentation, also known as data augmentation, is essentially the process of generating incremental data based on existing limited data without actually collecting more data, thereby creating value equivalent to a larger dataset. The essence of data augmentation methods lies not only in increasing the quantity of data samples but also in enhancing the features of the data itself. Sample data is a sampling of the entire dataset, and when the sample data size is large enough, the distribution of the samples should be similar to that of the overall population. However, due to objective reasons, the collected sample data may not be complete enough. In such cases, data augmentation methods can be used to generate new samples of data that are more similar to the distribution of real data. Deep learning neural network models possess strong learning capabilities, and thus, learning some useless information features can have a negative impact on the final results. Data augmentation techniques can impose constraints on the data according to requirements to increase the prior knowledge, such as deleting or supplementing some information, in order to reduce the negative impact on the model performance of processing image tasks.

Data augmentation methods are mainly divided into two categories: those applied to images and those applied to text. This paper primarily focuses on image-based data augmentation methods [1]. Based on whether machine learning techniques are used, these methods are divided into two parts for discussion: image data augmentation based on traditional image processing techniques and image data augmentation based on machine learning techniques. In the section on image data augmentation based on traditional image processing techniques, geometric transformations, color transformations, and pixel transformations targeting image data itself will be introduced. The section on image data augmentation based on machine learning techniques will cover automatic data augmentation techniques, data augmentation techniques based on generative adversarial networks, and data augmentation methods combining au-to-encoders and generative adversarial networks. These methods utilize machine learning-related theories to implement image generation and image transformation models. The image data augmentation method is shown in Fig 1.

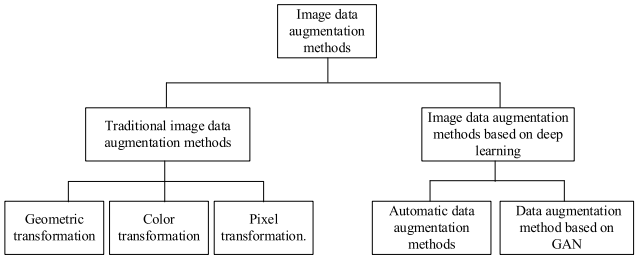


FIGURE 1. Classification of image data augmentation methods.

B. TRADITIONAL IMAGE DATA AUGMENTATION METHODS

Traditional image data augmentation methods typically utilize image processing techniques [2] to expand the dataset and optimize image quality, generally divided into three main categories: geometric transformations, color transformations, and pixel transformations.

Geometric transformations involve operations such as image flipping and rotation, image cropping and resizing, and image shifting and edge padding. Image flip-ping operations include vertical and horizontal flipping of images. Vertical

flipping is achieved by first performing horizontal flipping followed by a 180-degree rotation of the image. Horizontal flipping is more commonly used than vertical flipping. Whether the dimensions of the resulting image after rotation are the same as the original image depends on the degree of rotation and the shape of the original image. When a rectangular image is rotated 180 degrees or a square image is rotated 90, 180, or 270 degrees, the dimensions of the rotated image remain consistent with those of the original image.

Random cropping of images can be viewed as randomly sampling from the original image and then restoring the sampled image data sample to the original image size. Image scaling can be divided into outward scaling and inward scaling. Unlike image cropping, outward scaling results in images larger than the original image, from which an image of the same size as the original image is cropped. Inward scaling, on the other hand, reduces the size of the original image and fills the areas beyond its boundaries in order to obtain an image of the same size as the original image.

Image shifting refers to moving the image along the horizontal and vertical axes without changing its size, and padding the edge parts of the image. After performing image shifting operations, the useful parts of most image data for image tasks will be located at the edges of the image. Therefore, during the training of computer vision tasks with deep learning models, the focus of attention will shift to arbitrary positions rather than just focusing on learning from the center region of the image. This operation can effectively improve the robustness of the model. The mathematical representation of image shifting can be expressed as follows:

$$I_{shifted}(x, y) = I(x + \Delta x, y + \Delta y) \tag{1}$$

where Δx and Δy represent the shifts along the x and y axes respectively.

After performing operations such as rotation, shifting, and scaling on image data samples, it is necessary to restore the transformed images to the same size as the original image. The restoration process is achieved by padding the edge parts of the image. Commonly used image padding methods include: constant padding, which fills the edge parts of the image with constant values. This padding method is suitable for images with a single-color back-ground; boundary value padding, which fills the outside of the original image boundary with the pixel values of the original image boundary. This method is suitable for short-distance shifts.

Color transformations involve converting between different color spaces. Digital image data is represented using dimensions for width, height, and channels. Common color spaces include RGB color space, YUV color space, and HSV color space, among others. Color space conversion is a highly effective way to extract color features. Although different color spaces have their own characteristics, they can be converted between each other due to their isomorphic nature. For instance, Lu et al. [3] proposed a color space framework for facial recognition tasks, introducing the LuC1C2 color space. This color space selects the Lu luminance component by comparing the color sensor attributes of RGB coefficients. The direction of the transformation vector for the C1C2 color components is determined through the chromatic subspace of the RGB color space and covariance analysis. Experimental results conducted on facial image data-bases such as AR, Georgia Tech, FRGC, and LFW demonstrated that the LuC1C2 color space exhibits superior facial recognition performance.

Pixel transformations include noise, blur, and image fusion. Image noise refers to randomly superimposing isolated pixels or pixel blocks on the original image to disrupt observable information, thereby improving the generalization ability of convolutional neural network models. Common types of noise include salt-and-pepper noise and Gaussian noise. Blurring essentially involves convolving the original image, with Gaussian blur being a commonly used method. This method employs a convolutional kernel matrix that follows a two-dimensional normal distribution to reduce differences in pixel values, thus smoothing out the pixels of the image and achieving the effect of blurring the image. The Gaussian blur can be expressed as:

$$I_{blurred}(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G(i, j) \cdot I(x + i, y + j) \tag{2}$$

Image fusion techniques blend two images together by averaging the pixel values of the two images or by randomly cropping and stitching images together to form a new image. Better results can be obtained when blending images from the entire training set rather than just from instances of the same class. While image fusion methods may seem meaningless from a human perspective, they have been observed to improve accuracy in experiments. For example, CUTMIX is an improved random erasing strategy that randomly erases a portion of pixel information from the original image using a rectangular mask. However, its drawback is that it reduces the proportion of pixels containing information on the training image and requires a significant amount of computation, making it time-consuming.

C. DATA AUGMENTATION METHODS BASED ON DEEP LEARNING

In addition to traditional data augmentation techniques, researchers have begun to apply machine learning technologies to the field of data augmentation in recent years, achieving significant research outcomes. For instance, researchers from Google Brain proposed a method called Auto Augment [4], which automatically searches for suitable data augmentation policies. This approach designs data augmentation methods that do not alter the architecture of deep learning networks to achieve augmentation strategies with more invariance. This approach optimizes the training process of models from the perspective of policy search without modifying the neural network architecture. The method creates a search space to store data augmentation policies and selects appropriate sub-policies from the search space for

different batch tasks using a search algorithm. The selected sub-policies apply specific image processing functions for data augmentation operations, enabling the trained neural network to achieve the best validation accuracy. However, this method requires long training times under simplified settings.

Li et al. [5] proposed a new data augmentation technique, applying differentiable neural architecture search algorithms to data augmentation policy search tasks. This algorithm addresses the issue of expensive computation in Auto-Augment, which affects its applicability. The DADA algorithm transforms discrete data augmentation policy selection into an optimization problem using Gumbel-Softmax. CycleGAN [6] is an important model in the field of image transformation, enables the transformation of sample data without pairing, such as converting a celebrity into a cartoon character. This use of image transformation greatly expands sample data while retaining the contours of the original images. CycleGAN, as a method for unaligned data image transformation, is widely used for image-to-image transformation tasks. Yang He proposed a novel image generation method [7], classified as a stochastic regression approach, which learns to generate multiple different instances from a single conditional input. This method combines the advantages of generative adversarial networks and auto-encoders to accomplish image generation tasks, similar to the CVAE-GAN method. Chen et al. [8] introduced the Cascaded Refinement Networks (CRN), which transform image generation tasks into regression problems. This model demonstrates that it can synthesize image data seamlessly scaled to high resolution using a properly structured feed-forward network and proves the effectiveness of the model in experiments.

D. E-COMMERCE IMAGE DATA AUGMENTATION

The visual appeal of e-commerce product images plays a crucial role in attracting and retaining customers. Some studies have utilized image enhancement techniques to improve product images. Traditional data augmentation methods include simple transformations, horizontal and vertical flips, scaling, and color adjustments [9]. However, given the specificity of e-commerce images, the above enhancement methods have the following shortcomings when applied to e-commerce image enhancement: simple transformations may not address complex product shapes and backgrounds; color adjustments may struggle to overcome lighting variations; scaling may lead to distortion, affecting visual quality; horizontal and vertical flips cannot resolve annotation errors or noise issues. With the rise of deep learning, Generative Adversarial Net-works (GANs) based on deep learning have gradually been applied to the field of im-age enhancement. However, GANs have high computational resource requirements, unstable image quality, and issues such as mode collapse. These problems will affect the practical effectiveness and application scope of e-commerce image enhancement.

Additionally, there are two main issues with e-commerce image datasets: class imbalance and poor generalization. Addressing the class imbalance issue, reference [10] proposes a dynamic balancing of positive and negative sampling gradients for each class, reducing class imbalance. However, its optimal performance requires fi-ne-tuning of hyperparameters. Furthermore, this algorithm cannot completely eliminate class imbalance in highly imbalanced datasets. Reference [11] effectively reduces competition between rare and common classes by grouping them into disjoint sets, but if classes do not naturally fall into well-defined groups, optimal performance may not be achieved. Techniques focused on rare classes are proposed in [12] and [13], but they require parameter tuning and may not guarantee compatibility with certain neural network architectures.

To address the generalization issue of data augmentation, reference proposed the MixUp technique [14], which generates new training samples by blending between two images. Specifically, it combines the pixel values of two images using weighted addition, while also averaging their labels with weights to produce new training samples. While MixUp can reduce overfitting and facilitate learning of boundaries between different classes, the blending process may lead to class confusion, and the changes to image content in generated samples are relatively minor, limiting the model's generalization ability. Reference [15] introduced the Copy-Paste data augmentation technique, which involves copying and pasting parts of an image from the same image into an-other image to create new training samples. This technique helps the model learn about objects from different positions and angles, thereby enhancing the model's robustness and generalization ability. However, it may introduce noise and lose important information, and the pasted parts may be highly similar to the original image, failing to provide sufficient sample variation, thus affecting the model's learning effectiveness. CutOut [16] covers or cuts out randomly selected rectangular regions in im-ages to generate new training samples. This technique helps the model learn robust-ness to local information in images, reducing sensitivity to noise and interference. However, random cutting may lead to the loss of important feature information, and the consistency between generated samples is low, limiting the model's generalization ability.

In summary, the methods discussed above have potential advantages in addressing class imbalance and increasing dataset diversity. However, their effectiveness may be influenced by the specific dataset and problem at hand, requiring targeted adjustments and settings to achieve optimal results. Additionally, for certain methods, there may be a trade-off between improving balance and addressing potential drawbacks, such as increasing training time or model complexity. In practical applications, it is necessary to consider various factors comprehensively and make decisions and optimizations based on specific circumstances.

This paper proposes a convolutional neural network-based image data augmentation method for e-commerce product

images. The method utilizes techniques such as image segmentation, background replacement, shadow generation, and logo addition, primarily targeting image recognition tasks in e-commerce product catalogs. Com-pared to traditional image enhancement methods, the proposed approach demonstrates stronger robustness as it enables neural networks to perceive product images under various backgrounds, lighting conditions, and multiple contextual settings. Moreover, it maintains a balance between training time and model complexity, offering significant advantages over existing methods.

III. PROPOSED METHOD

This paper proposes an algorithm for automatically generating enhanced images based on image segmentation. The algorithm consists of five key steps: image segmentation, primary color detection, background processing, shadow generation, and logo addition. As shown in Fig 2, the proposed e-commerce image enhancement algorithm comprises the following five essential steps. Firstly, utilizing image segmentation techniques, the complex content of the image is segmented into multiple regions, providing clear targets for subsequent processing. Subsequently, employing primary color determination methods, the system accurately extracts the primary colors from the image, aiding in maintaining overall consistency in subsequent processing. During the background processing stage, advanced image processing algorithms are applied to effectively remove or replace the background, thereby making the image more prominent and clearer. Following this, through shadow generation techniques, the system simulates light projection and shadow formation, imparting a sense of depth and realism to the image.

The algorithm takes the original RGB image as input, segments the image, deter-mines the primary color tone of the foreground image based on the segmented image, and then performs background deletion or replacement operations based on the determined RGB primary color to enhance the display effect of the foreground image. To further improve the visual effect of the image and enhance the robustness of the algorithm, shadow generation and logo addition operations are sequentially performed on the replaced background image, resulting in an image enhanced using this algorithm. Each step will be elaborated on in the following sections.

A. IMAGE SEGMENTATION

This paper utilizes the instance segmentation model FastSAM, its overview as shown in Fig 3. The proposed method involves two stages: All-instance Segmentation and Prompt-guided Selection [17]. The first stage serves as the foundation, while the second stage acts as task-specific post-processing. Unlike end-to-end transformer models [18], this approach integrates various human priors that align well with vision segmentation tasks, such as convolutional local connections and strategies for assigning objects based on receptive fields. This customization enhances its suitability for vision

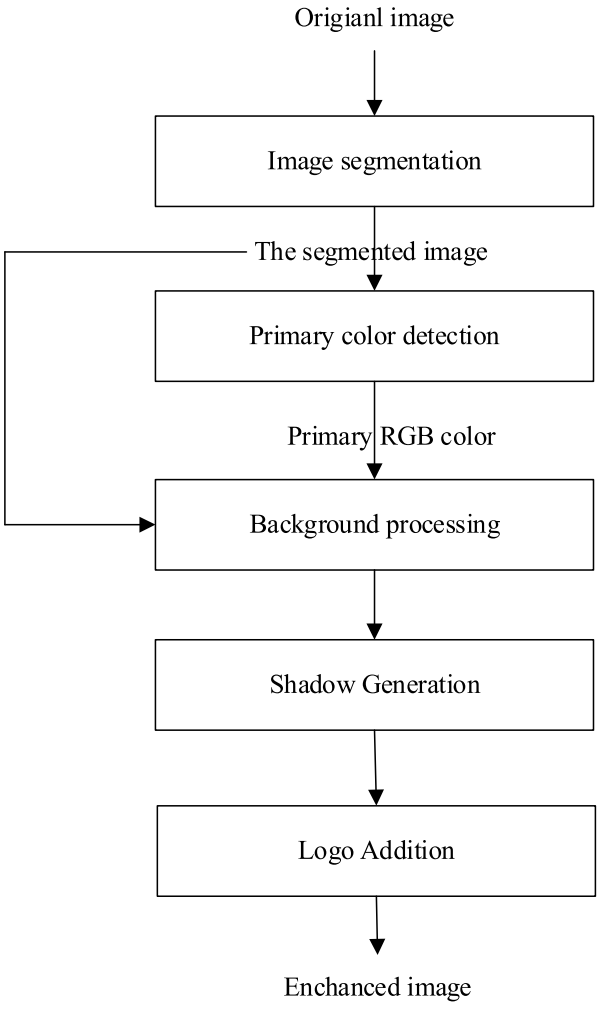


FIGURE 2. The workflow of the image enhancement algorithm in these papers.

segmentation, allowing for quicker convergence with fewer parameters.

1) ALL-INSTANCE SEGMENTATION

Model Architecture. YOLOv8 builds upon the foundation of its predecessor, YOLOv5, while incorporating significant design elements from more recent models, including YOLOX, YOLOv6, and YOLOv7. In YOLOv8, the backbone network and neck module replace the C3 module used in YOLOv5 with the C2f module. Additionally, the *Head* module has been enhanced with a decoupled architecture, separating the classification and detection heads, and transitioning from an Anchor-Based to an Anchor-Free approach.

Instance Segmentation: YOLOv8-seg leverages the principles established by YOLACT [19] for performing instance segmentation. The process begins by extracting features from an image using a backbone network combined with a Feature Pyramid Network (FPN) [20], which merges features of varying scales. The network's output is divided into

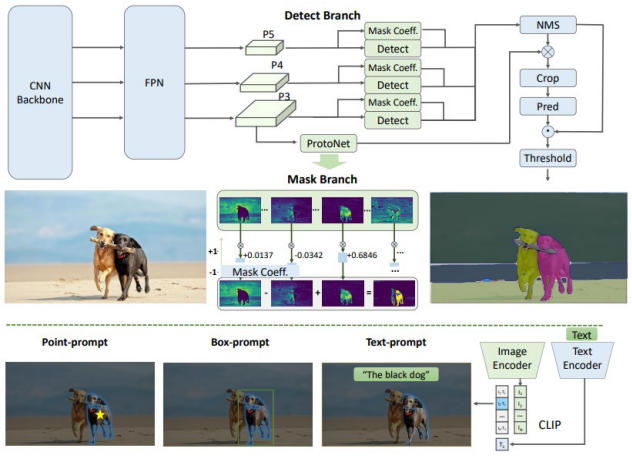


FIGURE 3. The framework of FastSAM. It contains two stages: All-instance Segmentation (AIS) and Prompt-guided Selection (PGS).

two branches: detection and segmentation. The detection branch produces the object categories and their corresponding bounding boxes, while the segmentation branch generates k prototypes (with 32 as the default in Fast SAM) along with k mask coefficients. Both tasks, detection and segmentation, are carried out simultaneously.

The segmentation branch receives a high-resolution feature map that retains spatial details while also encoding semantic information. This map undergoes processing through a convolutional layer, is then upscaled, and passes through two additional convolutional layers to produce the masks. The mask coefficients, similar to those in the detection head's classification branch, are within the range of -1 to 1. The final instance segmentation result is achieved by multiplying these mask coefficients with the prototypes and summing the results.

YOLOv8 is adaptable for a wide range of object detection tasks. With the addition of the instance segmentation branch, YOLOv8-seg becomes well-suited for the ''segment anything'' task, which focuses on precisely identifying and segmenting every object or region within an image, independent of the object category. The prototypes and mask coefficients offer significant flexibility for prompt-based guidance. For instance, a straightforward prompt encoder-decoder structure can be trained, where various prompts and image feature embeddings serve as inputs, and mask coefficients are produced as outputs. In Fast SAM, the YOLOv8-seg method is directly employed for the comprehensive instance segmentation stage. Although further refinement through manual design might yield additional improvements, such exploration falls beyond the scope of this work and is left for future research.

2) PROMPT-GUIDED SELECTION

After successfully segmenting all objects or regions within an image using YOLOv8, the subsequent phase in the segment anything task focuses on identifying the specific object(s) of interest through the use of various prompts. This process primarily involves point prompts, box prompts, and text prompts.

a: POINT PROMPT

This technique involves selecting specific foreground and background points in the image. Foreground points identify relevant masks, while background points exclude irrelevant ones. Morphological operations, such as dilation and erosion, are applied to enhance the precision of the final mask. This method is particularly effective for isolating irregularly shaped objects.

b: BOX PROMPT

Box prompts use a bounding box to guide the selection process. The Intersection over Union (IoU) is calculated between the selected box and the bounding boxes of the generated masks. The mask with the highest IoU score is chosen as the target object. This approach is well-suited for objects with defined geometric boundaries.

c: TEXT PROMPT

Using CLIP-based embeddings, text prompts allow semantic matching between textual descriptions and segmented regions. The text embeddings are compared with image features, and the mask most similar to the text embedding is selected. This method introduces a semantic dimension, enabling the model to process natural language inputs and refine segmentation accordingly.

The prompt-guided selection process builds on the results from the all-instance segmentation stage. While the segmentation stage ensures that all objects in the image are identified, prompt-guided selection enables targeted refinement to extract specific objects of interest. This integration makes the system versatile and capable of handling a wide range of use cases. For example, point and box prompts are ideal for applications requiring quick and precise object selection, while text prompts add semantic understanding for more complex scenarios.

By combining all-instance segmentation with prompt-guided selection, the proposed method achieves both high precision and flexibility. As depicted in Figure 3, this workflow enables the system to handle complex segmentation tasks efficiently, making it a robust solution for diverse e-commerce and real-world applications.

B. PRIMARY COLOR DETECTION

After obtaining the segmented image, the next crucial step is to determine the primary color of the final object. This paper employs the unsupervised learning algorithm k-means [19] to achieve primary color determination. The K-means algorithm aims to cluster image data into K clusters, with the core idea of minimizing the total squared deviation between data points and their cluster centroids. Through an iterative process, the K-means algorithm determines the cluster centroids of data points, thereby identifying the primary colors in the image.

In practical applications, by selecting an appropriate value for K, it is possible to ensure that the centroid of the largest cluster accurately represents the main colors of the image, thus extracting the primary colors of the image.

The primary color detection step provides a critical foundation for subsequent image processing steps, including background replacement and shadow generation, ensuring that the overall consistency and visual appeal of the image are maintained throughout the workflow. In the background replacement stage, the identified primary color guides the selection of complementary or harmonious background colors, ensuring that the product remains visually prominent while maintaining aesthetic coherence. For instance, products with a dominant blue tone are paired with neutral or cool-toned backgrounds to enhance visual harmony, whereas products with warm tones, such as red or orange, are paired with softer, contrasting backgrounds to avoid clashing colors.

Similarly, in the shadow generation stage, the primary color is used to derive the shadow tones, creating shadows that blend seamlessly with the object and its surrounding environment. Shadows are generated by applying a darker shade of the detected primary color, which prevents artificial visual discrepancies and maintains the natural appearance of the product.

C. BACKGROUND PROCESSING

The background processing module plays a crucial role in enhancing the visual appeal and clarity of e-commerce product images by either removing or replacing the original background. This module offers three distinct approaches to background manipulation: 1) solid color fill, 2) texture replacement, and 3) image composition. Each approach is designed to cater to different aesthetic and functional requirements.

1) SOLID COLOR FILL

In this approach, the original background is replaced with a uniform color, such as white or a gradient. This method is simple yet effective, as it ensures that the product remains the focal point of the image. It is particularly suitable for platforms that prioritize clean and minimalistic designs. Mathematically, the operation can be expressed as:

$$I_{output}(x,y) = \begin{cases} I_{foreground}(x,y) & if\ M(x,y) = 1 \\ C_{background} & if\ M(x,y) = 0 \end{cases} \quad (3)$$

where $I_{foreground}$ represents the product image, $C_{background}$ is the chosen solid color, and $M(x,y)$ is the binary mask generated during instance segmentation.

2) TEXTURE REPLACEMENT

This technique overlays the segmented product onto a predefined textured background, adding richness and depth to the visual presentation. The texture is selected from a library of patterns tailored to the product category, ensuring harmony with the product's primary color. The blending operation uses alpha compositing:

$$I_{output} = \alpha I_{foreground} + (1-\alpha)\, T_{backgroud} \quad (4)$$

where α is the blending coefficient, and $T_{backgroud}$ represents the textured background.

3) Image Composition

In this method, the foreground object is seamlessly integrated onto a new photographic background, creating a natural and realistic environment for the product. The integration preserves the object's edges and details using Alpha blending, as defined by:

$$I_{output}(x,y) = M(x,y) \cdot I_{foreground}(x,y) + (1 - M(x,y)) \cdot I_{backgroud}(x,y) \quad (5)$$

Here, $I_{backgroud}$ denotes the new background image, and $M(x,y)$ is the binary mask indicating the object's position.

By offering these three approaches, the background processing module provides flexibility in designing product images for different e-commerce scenarios. This ensures that the enhanced images not only meet aesthetic standards but also highlight the unique features of the products effectively.

D. SHADOW GENERATION

The shadow generation module is essential for enhancing the three-dimensional appearance and realism of e-commerce product images. By simulating the effects of light projection, this module creates shadows that mimic real-world lighting conditions, thereby making the product visually compelling. The process begins by determining the light source direction and intensity, which influence the shape, size, and opacity of the shadow. The shadow is then generated using a combination of geometric transformations and blending techniques. The mathematical representation of shadow generation can be expressed as:

$$S(x',y') = \alpha \cdot F(x,y) + (1-\alpha) \cdot B(x',y') \quad (6)$$

where $S(x',y')$ represent the pixel value of the shadow-augmented image at position (x',y'), $F(x,y)$ denotes the pixel value of the foreground product, $B(x',y')$ refers to the pixel value of the background, and α is the shadow intensity coefficient, which ranges for 0 to 1.A higher α value result in a darker shadow. To ensure natural blending, Gaussian blur is applied to soften shadow edges. The Gaussian kernel, defined as:

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (7)$$

where $G(x,y)$ is the kernel value at position (x,y),and σ represents the standard deviation, controlling the blur radius and smoothness of shadow transitions. This combination of shadow modeling and blending ensures the enhanced image maintains high aesthetic quality and visual coherence.

To further enhance the realism of the shadows, the proposed method incorporates specular shadow generation, which simulates light reflections on the product surface.

Specular shadows add dynamic highlights and depth, mimicking real-world lighting effects. This is achieved using a modified Blinn-Phong reflection model, which calculates the intensity of light reflections as:

$$I = k_d \cdot (L \cdot N) + k_s \cdot (R \cdot V)^n \tag{8}$$

In the modified Blinn-Phong reflection model, the intensity of the reflected light I is calculated by combining diffuse and specular reflection components. The diffuse reflection component$k_d \cdot (L \cdot N)$, represents the scattered light intensity on the surface. Here, k_d is the diffuse reflection coefficient, controlling the proportion of light scattered uniformly, while L and N denote the light source direction and surface normal vector, respectively, with their dot product determining how directly the light strikes the surface. The specular reflection component, $k_s \cdot (R \cdot V)^n$, simulates light reflections concentrated around the viewer's perspective. In this term, k_s is the specular reflection coefficient, which dictates the intensity of the specular highlight, R represents the direction of the reflected light, and V denotes the viewer direction. The shininess factor n controls the sharpness of the highlight, with higher values producing smaller, more concentrated reflections typical of polished surfaces. Together, these parameters enable the dynamic adjustment of light reflection and shadow distribution, ensuring that the generated specular shadows align naturally with the product's material properties and surrounding lighting conditions.

Specular shadows are dynamically adjusted based on the product's material and texture. Glossy surfaces, such as metals or polished plastics, exhibit sharper and more intense highlights due to higher k_s values, while matte surfaces, like fabrics, produce softer and more diffused reflections. The direction and intensity of the light source are tailored for each product, ensuring that the shadows integrate naturally with the background and surrounding elements.

E. LOGO ADDITION

Logos are not just decorations, which they can also contain additional information about the image content, such as brand names, product models, website links, etc. This additional information provides more context to the image, making it easier for image recognition algorithms to understand the content conveyed by the image. In this step, we take the RGB image with replaced background and generated shadows, along with the logo, as input. By combining the image with the logo in an overlapping manner, we generate high-quality enhanced images. This step not only makes the image more personalized and specialized but also adds brand identification or other relevant information to enhance the commercial value and recognition ability of the image. By adding the logo, we can add more information and meaning to the image, making it stand out more in conveying information and attracting attention.

To maintain visual coherence and avoid interfering with key product features, the placement and size of the logo are dynamically determined. The logo is positioned in an unobtrusive yet prominent area, such as the bottom-right or top-left corner of the image. Bounding box coordinates generated during the segmentation step are analyzed to identify free background space, ensuring the logo does not overlap with essential product details. Additionally, the size of the logo is scaled proportionally to the dimensions of the product's bounding box, typically occupying 5–10% of the area. This approach ensures that the logo remains visually noticeable without detracting from the product's prominence in the image.

By combining the image with the logo in this manner, we generate high-quality enhanced images that are personalized and specialized. These logos not only provide brand identification or other relevant information but also enhance the commercial value and recognition ability of the image. For example, Figure 7 demonstrates two cases where logos are dynamically added to product images. In the first case, a footwear product is paired with a logo positioned in the bottom-right corner, scaled to 7% of the product's bounding box area. In the second case, a fashion accessory features a logo in the top-left corner, occupying 5% of the product's bounding box area. These examples highlight the adaptability of the logo addition process across different product types.

By adding logos in this structured manner, the processed images gain more contextual meaning and become more visually appealing, standing out in conveying information and attracting attention. This step ensures that the enhanced images not only meet the aesthetic and informational needs of e-commerce applications but also provide added branding value for commercial purposes.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL PLATFORM AND DATASET ANALYSIS

The hardware platform used in this experiment is the Dell PowerEdge T640, con-figured with an Intel Xeon Gold 6226R CPU, featuring 16 cores and a clock speed of 3.22GHz. It is equipped with an NVIDIA GeForce RTX 3080 GPU with 8GB of VRAM and 16GB of RAM. The operating system is Ubuntu 20.04 LTS. The software platform includes PyCharm 2022.2, Python 3.8, PyTorch 1.7, CUDA Toolkit 11.0, and cuDNN 8.0.

The experiments in this study utilized a comprehensive e-commerce image dataset specifically curated for evaluating image enhancement methods. The dataset consists of 10 product categories, including clothing, footwear, accessories, and electronic devices, carefully selected to represent the diversity of products commonly sold on e-commerce platforms. These categories were chosen to capture a wide range of challenges associated with e-commerce image processing, such as the need for accurate color representation in clothing, texture differentiation in footwear, and shape recognition in electronics. By focusing on these representative categories, the dataset aligns closely with real-world e-commerce applications.

A total of 2,618 images were collected from publicly available online sources, ensuring a variety of shooting conditions,

such as different lighting setups (indoor, outdoor, and low-light environments) and background types (plain, textured, and natural). Approximately 40% of the images feature complex natural backgrounds, 35% have plain backgrounds, and 25% include textured or gradient backgrounds, reflecting the visual diversity encountered in e-commerce scenarios. The dataset also includes products of various colors, shapes, and sizes to ensure robust evaluation across heterogeneous data. For instance, clothing images feature both single-colored and patterned items, while electronic devices vary in form factor and surface finish.

Each image in the dataset was annotated with key attributes, such as product category, brand, color, and size, to support downstream tasks and enhance its applicability. Prior to training and evaluation, all images were preprocessed, including resizing to a uniform resolution of 640 × 640, normalization, and label encoding, ensuring consistency and compatibility with the proposed method. Examples of the dataset images are shown in Figure 4, demonstrating the diversity and complexity of the data.

FIGURE 4. Some images from the e-commerce image dataset.

Despite its strengths, the dataset has certain limitations that may affect the generalizability of the results. The focus on commonly sold product types means that niche categories, such as furniture or automotive parts, are underrepresented. Additionally, the dataset's size, while diverse, may not fully capture the vast range of product variations encountered in large-scale e-commerce platforms. These limitations suggest that while the proposed method demonstrates robust performance on the included categories, future work involving larger datasets with broader category representations would be necessary to further validate its generalizability. By addressing these limitations and emphasizing its strengths, this dataset provides a valuable benchmark for evaluating image enhancement techniques while offering insights into its applicability in real-world e-commerce scenarios.

B. ALGORITHM PERFORMANCE

Based on the dataset constructed in the previous sections, running the algorithm proposed in this paper enables the automatic generation of images that comply with the standards of the e-commerce industry. Examples of enhanced images are shown in Figure 5. In this example, we start with a light gray background and process an im-age of a sports wristband. Firstly, the algorithm automatically segments the image and replaces the background with a solid color or a specific scene, making the wristband image more prominent and clearer. Secondly, by generating specular shadows, the three-dimensional and realistic appearance of the image is enhanced, making the wristband look more lifelike. Finally, the logo addition feature ensures the reasonable calculation of the logo's position, avoiding obstruction of objects and maintaining the overall aesthetics of the image. Overall, the images processed by the algorithm in this paper exhibit superior visual effects and meet the standard requirements of the e-commerce industry.

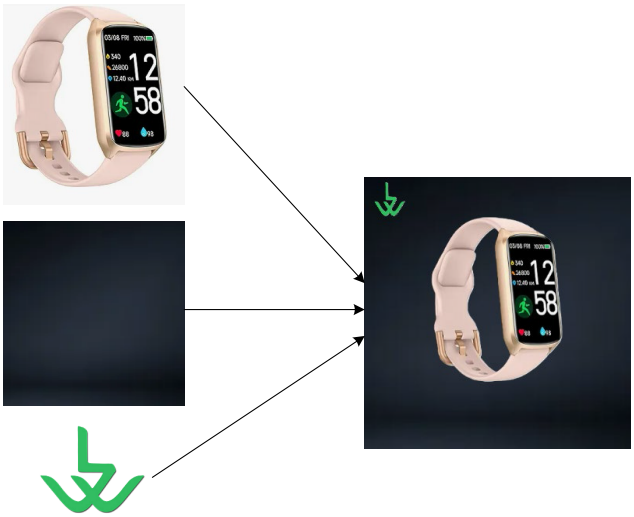


FIGURE 5. Example of enhanced image generated by the proposed algorithm.

C. VALIDATION OF EFFECTIVENESS

To further validate the effectiveness of our algorithm, we conducted tests to determine whether image enhancement contributes to improving automatic image recognition. Firstly, we defined the dataset constructed earlier as the base dataset, which includes the original images along with their corresponding labels. Next, we ap-plied our proposed algorithm to enhance each image in the base dataset, generating the enhanced dataset II. Then, we randomly selected 50% of the images from both the base dataset and enhanced dataset II to form enhanced dataset I. In this way, we obtained three different datasets: the base dataset, enhanced dataset I, and enhanced dataset II.

Subsequently, we divided these three datasets into training and testing sets and utilized the YOLOv5, SSD and Faster RCNN image recognition model for training and testing. On the testing set, we evaluated the segmentation performance and conducted comparative analysis on metrics such as accuracy, IoU (Intersection over Union), average precision. The formulas for Accuracy, IoU and mAP are as follows:

$$IOU = \frac{A \cap B}{A \cup B} \tag{9}$$

where A is the marking window, B is the detection window, the numerator represents the overlapping area of windows A

TABLE 1. The comparison of YOLOv5, SSD and FASTER-RCNN performance on the three datasets.

Algorithm	Dataset	Accuracy	IoU	mAP
YOLOv5	Basic Dataset	0.78	0.73	0.74
	Enhanced Dataset I	0.85	0.82	0.84
	Enhanced Dataset II	0.91	0.90	0.92
SSD	Basic Dataset	0.84	0.81	0.85
	Enhanced Dataset I	0.90	0.85	0.89
	Enhanced Dataset II	0.94	0.92	0.95
Faster-RCNN	Basic Dataset	0.88	0.87	0.87
	Enhanced Dataset I	0.92	0.91	0.93
	Enhanced Dataset II	0.96	0.94	0.95

and B, and the denominator represents the sum of the areas of windows A and B. Obviously, the value of IOU is between [0, 1]. The closer the IOU is to 1, the more the two windows overlap, and the better the positioning accuracy. Otherwise, the worse it is.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

Among them:

True positives (TP): the number of instances correctly classified as positive, that is, the number of instances that are actually positive and classified as positive by the classifier Q (number of samples).

True negatives (TN): the number of instances correctly classified as negative, that is, the number of instances that are actually negative and classified as negative by the classifier.

False positives (FP): the number of instances incorrectly classified as positive, that is, the number of instances that are actually negative but classified as positive by the classifier,

False negatives (FN): the number of instances incorrectly classified as negative, that is, the number of instances that are actually positive but classified as negative by the classifier.

The specific results are presented in Table 1. This experiment aims to explore the impact of our algorithm on automatic image recognition tasks, validating the potential of image enhancement techniques in improving image processing and recognition performance, and providing strong empirical support for further research in image processing.

The experimental results indicate that the proposed data augmentation method, when applied to the YOLOv5, SSD, Faster-RCNN model for e-commerce image recognition tasks, has a significant positive impact on recognition accuracy, IoU and mAP. The image dataset processed with data augmentation during the training process effectively improves the model's generalization ability and facilitates the model to converge to higher performance levels in a shorter time. This result further validates the rationale behind the additional cost of longer training times during training, as it significantly enhances the model's recognition performance and robustness. The findings of this experiment provide strong support for employing data augmentation techniques to improve the performance of deep learning models in e-commerce image recognition tasks. They also offer important reference value for further research and application of image data augmentation methods.

To further validate the effectiveness of the proposed method, we extended our experimental analysis by including comparisons with additional methods. Specifically, we selected representative techniques from both traditional and advanced image enhancement approaches. Table 2 presents the comparative performance of these methods against our proposed algorithm on the e-commerce dataset.

TABLE 2. Performance comparison of the proposed method and baseline techniques.

Method	Accuracy	IoU	mAP	Time per Image (ms)
Geometric Transformations	0.82	0.78	0.79	110
Color Adjustments	0.83	0.79	0.80	120
CycleGAN	0.89	0.86	0.87	280
Proposed Method	**0.91**	**0.90**	**0.92**	**125**

The results in Table 2 demonstrate that the proposed method outperforms traditional data augmentation techniques and CycleGAN across all performance metrics. Specifically, it achieves the highest accuracy (91.0%), IoU (90.0%), and mAP (92.0%) while maintaining a competitive average processing time of 125 ms per image. In comparison, CycleGAN, while performing well in recognition metrics, has a significantly higher computational cost of 280 ms per image, making it less suitable for real-time applications. The proposed method effectively balances performance and efficiency, highlighting its potential for practical deployment in e-commerce platforms.

Additionally, to further validate the necessity of the segmentation algorithm, we conducted an ablation study comparing the performance of the proposed method with

TABLE 3. Performance comparison with and without the segmentation algorithm.

Method	Accuracy	IoU	mAP	Time per Image (ms)
Without Segmentation Algorithm	0.85	0.82	0.83	100
With Segmentation Algorithm	**0.91**	**0.90**	**0.92**	**125**

and without the segmentation step. Table 3 summarizes the results.

The results demonstrate that incorporating the segmentation algorithm significantly improves performance across all metrics. The accuracy increased by 5.8%, IoU by 7.7%, and mAP by 8.5% when the segmentation algorithm was used. Although the inclusion of segmentation added an additional processing cost of 25 ms per image, the performance gains justify its integration into the workflow. These findings highlight the segmentation algorithm's critical role in reducing noise from complex backgrounds, improving the clarity of input data, and enabling more accurate downstream processing.

While the performance metrics reported in this study demonstrate significant improvements across models, statistical validation such as confidence intervals or t-tests was not conducted due to the scope of the current study. The primary focus was to evaluate the effectiveness of the proposed method through comparative metrics such as accuracy, IoU, and mAP. Nevertheless, future work will aim to incorporate statistical analyses to further validate the reliability and generalizability of the results.

V. DISCUSSION

This study highlights the substantial performance gains achieved by the proposed method across different models and metrics, demonstrating its potential impact in both technical and practical contexts. Beyond achieving significant improvements in accuracy (91.0%), IoU (90.0%), and mAP (92.0%) when applied to YOLOv5, SSD, and Faster-RCNN, the proposed method offers broader implications for e-commerce applications and other domains.

The proposed method's ability to enhance image quality has direct relevance to e-commerce metrics such as click-through rates (CTR) and conversion rates (CR). High-quality images, characterized by improved segmentation, realistic shadows, and seamless logo integration, are more visually appealing and can capture user attention effectively. This is likely to result in higher CTR as users are drawn to better-presented product images. Similarly, the increased accuracy of image recognition models can enhance recommendation system precision and search engine relevance, thereby driving higher CR. For example, products with accurately enhanced images are more likely to appear in relevant searches, improving the overall user experience and increasing purchase likelihood.

The technical features of the proposed method, including instance segmentation, background replacement, and shadow generation, offer promising applications in fields such as natural landscape photography, medical imaging, and artwork analysis. In natural landscape photography, these techniques can isolate elements like trees, mountains, and water bodies for targeted enhancements such as color balancing or sky replacement. In medical imaging, precise segmentation supports the isolation of anatomical structures, aiding diagnostics and improving visualization. Similarly, in artwork analysis, the method's ability to manage complex backgrounds and detect dominant colors can facilitate restoration, study, and high-fidelity digital replication of artworks.

While the method demonstrates robust performance, the lack of statistical validation, such as confidence intervals or significance tests, represents a limitation. This restricts the ability to fully quantify the reliability of performance improvements. Additionally, the dataset used in this study, while diverse, primarily focuses on common e-commerce product categories and may not fully represent niche categories. These limitations suggest that further work is required to validate the generalizability of the findings and address broader datasets and scenarios.

To address these identified limitations, future research will incorporate rigorous statistical analyses, including confidence intervals and significance tests, to ensure a more robust foundation for the reported improvements. Expanding the dataset to include more niche product categories and testing the method's scalability on larger, more heterogeneous datasets will also be prioritized. Additionally, optimizing the algorithm for domain-specific applications, such as adaptive segmentation for medical imaging or artistic style transfer for artwork analysis, offers exciting avenues for further exploration.

VI. CONCLUSION

This paper presents an innovative e-commerce image enhancement technique that integrates instance segmentation, primary color detection, background replacement, shadow generation, and logo addition into a unified framework. The proposed method achieves significant improvements in accuracy, IoU, and mAP while maintaining computational efficiency, with an average processing time of 125 ms per image. These results highlight the method's suitability for real-time e-commerce applications and its potential to enhance image quality and recognition accuracy across various scenarios.

By aligning technical capabilities with the unique demands of different fields, this study demonstrates the versatility and scalability of the proposed method. The findings lay a foundation for future work in advancing image processing methodologies and expanding the algorithm's applicability across diverse industries, from e-commerce to medical imaging and beyond.

REFERENCES

[1] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Netw. Vis. Recognit.*, vol. 11, pp. 1–8, Jan. 2017.

[2] H. Zhu, *Fundamentals of Digital Image Processing*. Beijing, China: Science Press, 2005.

[3] Z. Lu, X. Jiang, and A. Kot, "Enhance deep learning performance in face recognition," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 244–248.

[4] E. Cubuk, B. Zoph B, D. Mane, V. Vasudevan, and Q. Le, "Auto augment: Learning augmentation policies from data," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 113–123.

[5] Y. Yi, G. Hu, Y. Wang, T. T. Hospedales, N. Robertson, and Y. Yang, "DADA: Differentiable automatic data augmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 580–595.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[7] Y. He, B. Schiele, and M. Fritz, "Diverse conditional image generation by stochastic regression with latent drop-out codes," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 406–421.

[8] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1520–1529.

[9] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[10] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, and D. Lin, "Seesaw loss for long-tailed instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9690–9699.

[11] Y. Li, T. Wang, B. Kang, S. Tang, C. Wang, J. Li, and J. Feng, "Overcoming classifier imbalance for long-tail object detection with balanced group softmax," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10988–10997.

[12] J. Tan, X. Lu, G. Zhang, C. Yin, and Q. Li, "Equalization loss v2: A new gradient balance approach for long-tailed object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1685–1694.

[13] C. Esposito, G. A. Landrum, N. Schneider, N. Stiefl, and S. Riniker, "GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning," *J. Chem. Inf. Model.*, vol. 61, no. 6, pp. 2623–2640, Jun. 2021.

[14] Y. Chen, V. T. Hu, E. Gavves, T. Mensink, P. Mettes, P. Yang, and C. G. M. Snoek, "PointMixup: Augmentation for point clouds," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 330–345.

[15] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2917–2927.

[16] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.

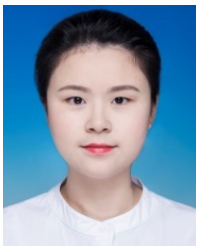
[17] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, "Fast segment anything," 2023, *arXiv:2306.12156*.

[18] B. Cheng, A. Schwing, and A. Kirillov, "Perpixel classification is not all you need for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Systems.*, vol. 34, 2021, pp. 17864–17875.

[19] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9156–9165.

[20] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

QIANG GAO received the master's degree from South China Normal University, in 2010. He is currently an Associate Professor with the School of Information Engineering, Guangzhou Railway Polytechnic, Guangzhou, China. He has published several articles, such as *Computer System Applications* and *Computer Technology and Development*. His major research interests include pattern recognition, image recognition, deep learning, and image enhancement.



HUIPING HU received the Ph.D. degree in educational management from the Graduate School, St. Paul University Philippines. She is currently an Associate Professor with the Department of Primary Education, Shangrao Preschool Education College, Shangrao, China. Her major research interests include English education, vocational English teaching, and language deep learning. She has published several articles in such journals as *Applied Mathematics*, *Nonlinear Science*, and other Chinese journals.

WEI LIU received the master's degree from South China Normal University, in 2008. She is currently an Associate Professor with the School of Information Engineering, Guangzhou Railway Polytechnic, Guangzhou, China. Her major research interests include artificial intelligence and big data technology.