Studies in Media and Communication
Vol. 11, No. 6; September 2023
ISSN: 2325-8071 E-ISSN: 2325-808X
Published by Redfame Publishing
URL: http://smc.redfame.com

AR Technology-Assisted Selfie Design and Dissemination Path Analysis

Di Zhang[1], Syed Agil Alsagoff[1], Megat Al Imran Yasin[1], Siti Aishah Muhammad Razi[1]

[1]Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Selangor,43400, Malaysia

Correspondence: Di Zhang, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Selangor,43400, Malaysia.

Received: June 26, 2023 Accepted: July 29, 2023 Online Published: July 30, 2023

doi:10.11114/smc.v11i6.6203 URL: https://doi.org/10.11114/smc.v11i6.6203

Abstract

Augmented reality (AR) is a technology based on 3D registration, virtual-real fusion and human-computer interaction to achieve the integration of virtual objects and real scenes. The core problem of augmented reality task is the accurate and fast recognition and tracking of objects in real scenes, which provides the technical basis for updating and optimizing the selfie design. This paper discusses the specific technical path of AR applied to selfie design based on the deep learning approach, and demonstrates the impact of different deep learning algorithms on the effectiveness of the integration of AR and selfie, and finally the paper discusses the application prospects of AR in selfie.

Keywords: augmented reality, selfie design, deep learning approach, dissemination path

1. Introduction

With the expansion of the "watershed" in the Internet age, especially the development of digital photography technology and social media, it has provided a technical and sharing platform for selfie-takers, making "selfies" no longer the "exclusive" of artists and photographers, but part of the life of the general public. "In early 2019, the British media reported that approximately 93 million selfies are taken globally every day, and 880 billion selfies are shared online each year. It is clear that the "selfie" - the act of individuals taking photos with smartphones or webcams and sharing them through social media - has become a global craze that is growing in popularity - today, you can see people posing and gazing at selfies wherever you are, whether they are businessmen or businesswomen. They are either businessmen and celebrities, politicians and stars, or ordinary people, who share their selfies, whether documentary, beautiful, or quirky, on major social networking platforms for people to watch and comment (Wu et al., 2019). A survey in 2018 showed that the post-1985 group accounted for 81.8% of selfie users in China. Foreign survey data also shows that the average millennial will take more than 25,000 selfies in their lifetime. It can be argued that they are naturally connected to the Internet. It can be said that teenagers, who were born with new media, have formed an ideological consensus and lifestyle of "I take pictures, therefore I am", or that the selfie culture, which has flourished based on the Internet era, has become a subculture active among teenagers (Lyu et al., 2021).

Along with the rise of the Internet and digital technology came virtual reality (VR), a computer simulation technology that allows the creation and experience of virtual digital worlds with a high degree of immersion. Based on virtual reality technology, augmented reality (AR) has also been developed. Augmented reality (AR) is an interactive experience of real-world environments in which objects existing in the real world are augmented with computer-generated perceptual information, sometimes across multiple sensory modalities, including vision, hearing, touch, physical sensation, and smell (Zheng et al., 2018). This technology uses the virtual world on the screen to interact with the real world, thus merging reality with the virtual. With the help of computer graphics and visualization technologies, the user's senses are given a new experience in the real physical world, enhancing the user's perception of physical objects. Importantly, the development of AR provides a good basis for the combination of art and technology (Niu et al., 2019; Liu & Keane, 2020). The development of computer software has not only provided new forms of expression for art, but also unlimited creative inspiration for many artists and designers in the digital age, and personal selfie design is no exception. Applying AR technology to the design of selfies can effectively enhance the individual's self-identity and also improve the communication effect of selfies (Lu & Chia, 2022).

Therefore, this paper firstly reviews the development of AR technology and selfies respectively, secondly discusses the specific technical path of AR applied to selfies around the latest frontier development of AR, and briefly discusses the effectiveness of different models, and finally this paper discusses the application prospects of AR in the field of selfies.

2. Literature Review

Driven by technological changes, digital marketing communications have undergone a sea change. In particular, the development of mobile Internet technology has led to a dramatic shift in the form of scene marketing communication from real to virtual, from fixed to mobile, from offline to online, and from physical presence to mental absence. Among them, the emergence of AR technology provides a new path for user selfie design, and also has a great impact on the construction and communication of individual self-identity.

2.1 Definition of AR

Augmented reality is to generate virtual digital objects with the help of computer graphics and visualization technologies, and precisely nest the virtual objects in the real environment through sensing and tracking technologies, and then integrate the virtual objects with the real environment with the help of display devices to achieve real-time interaction, virtual-reality combination, and present a realistic experience for users to meet the sensory effects (He, 2019). The technical features of augmented reality include virtual reality fusion, real-time interaction and 3D registration. As a type of immersive experience, smooth real-time interaction is an important operational process for user perception, allowing users to experience more of it.

Nowadays, with the emergence of smart mobile devices such as cell phones, tablet PCs and holographic glasses, the foundation for the development of augmented reality technology has been laid, and in addition to the existing hardware devices, there are a large number of related applications and other software, and the public's expectations for augmented reality technology are gradually growing. Devices based on augmented reality can be broadly divided into two categories, one is head-mounted AR, and the other is AR based on mobile devices.

Compared with head-mounted AR devices, mobile phone-based AR devices are more popular because they are lighter and more widely used. Vuforia is the main development platform based on image recognition, which uses the device's camera to capture pictures in real space and calculate the location, and then places the corresponding virtual model on top of the pictures, which is similar to the QR code that users are more familiar with (Grubert et al., 2017). The second is based on SLAM spatial understanding. The second type of AR is based on SLAM spatial understanding, which is commonly used in robots at first to help them understand the internal environment of their work. When used on cell phones, it can calculate the physical space of reality (Mamone et al., 2020).

2.2 Key Technology of AR

Augmented reality is the fusion of virtual objects with reality, or the strengthening and improvement of the original environment, and the close connection between real life and virtual world, so as to achieve the effect of fusion of reality and reality. This kind of virtual reality can enhance the user's perception of real life and bring a new experience to the user over time and space, thus achieving the effect of augmentation. Therefore, AR has a wide range of application areas in selfie design, especially the core technology of AR has a natural fit with the selfie demand of consumers in the digital era. Specifically, the core technologies of AR include the following three aspects.

Object tracking technology is the decisive technology for augmented reality systems. The ability to correctly capture objects in the real environment and perfectly track their movements is a prerequisite for subsequent virtual object generation and display technologies, which can correctly track real entities, identify key features and display them in combination with virtual objects to obtain a better display effect. Neural networks and support vector machines (VSM) are used to further improve the image processing, identify and extract important features of entities, and facilitate the subsequent display of virtual objects in combination with real entities to highlight or improve the important features of real entities (Yang et al., 2022).

Virtual object generation technology is the core part of AR technology that distinguishes it from other image and video display solutions. Its role is to generate virtual objects that interact with real objects and display multi-dimensional information of real objects based on different human senses. The virtual object generation technology aims to generate more realistic virtual objects, which make full use of 3D modeling technology to reproduce every detail of real entities to form a perfect virtual world, so that these virtual objects can be deeply intertwined with the real world. At the same time, in the process of integration with the real world, the virtual objects need to be more softly and smoothly displayed dynamically, and can be transformed according to the user's operation, so as to display the characteristics of the objects in an all-round way, and make the user more comprehensive access to information content (Cao & Cerfolio, 2019).

Display technology is the core technology of AR technology that brings the most direct experience to users, and its role is to display the extended information of real entities and virtual information in multiple directions, which can show users more comprehensive and better content information. With the continuous advancement of technology, modern AR display technology has made great progress, and many large technology companies, such as Apple, Google, Microsoft, etc., have increased their investment in AR/VR technology and released a variety of AR/VR glasses, headgear and other display devices. These displays allow users to experience a virtual world that blends with the real world in a more immersive way, perfectly blending the virtual world with the real world and presenting it in a more colorful way. Reality technology not only serves as a display, but also includes some interactive interfaces that allow users to manipulate objects in the system according to their needs, further enriching the system functions of AR technology (Baran et al., 2019).

In sum, the iterations and updates of AR technology have enriched people's daily digital lives, but there is a lack of sufficient discussion on how AR can be applied to selfie design.

2.3 Selfie and AR Technology

The emergence of new media is a message in itself, not just a message conveyed by the media. This message has not only changed the spatial and temporal structure of information dissemination and expanded the depth and breadth of human communication, but also, and more importantly, it has brought about "technological empowerment". At this level, the selfie wave is mainly due to the growth of the new generation of the Internet and the empowerment of new media technology.

On the one hand, the growth of the new online generation has laid a solid user base for the development of selfies. If millennials, who witnessed the rise of search engines, mobile Internet and instant messaging, are the pioneers of digitalization, the Z-generation lived in a digital world even from birth: the Internet, smartphones, gaming devices, and social media ...... grew up with them, and they are the first batch of digital society They are the first "natives" of the digital society. According to some data, among the 854 million Internet users in China, the 10-39 age group accounts for 70.8% of the total Internet users. The White Paper on the Psychology and Behavior of Mobile Phone Use among Post-95s published by the School of Psychology and Cognitive Science of Peking University also shows that the average daily time spent by post-95s on cell phones is 8.33 hours (Aydoğdu, 2021). This shows that China's Internet users are mainly the growing new generation of Internet users, who have distinctive personalities, independent thinking and broad vision, and are eager to show themselves through selfies so as to obtain their identity.

On the other hand, the empowerment of youth by new media technologies has revolutionized the traditional mass media's centralized communication path, making selfies an everyday activity in which youth are widely engaged. The digital world offers the possibility for them to challenge social norms, explore fun, develop skills and experiment with self-expression." If the popularity of smartphones and the emergence of technologies such as high pixel, front-facing cameras, beauty cameras, and selfie sticks have expanded and enriched people's right and means to take selfies, the development of social media platforms has empowered people to disseminate and comment on selfies. In fact, most of the development of new media technologies is related to youth, and Internet technologies themselves are consistently linked to youth popular culture. Thus, it can be argued that new media technologies empower individuals to produce and disseminate information, thus empowering youth groups to "produce their own images" through the use of new media technologies (Braly et al., 2019). As a result, more and more selfie-takers are involved in the process of content production and dissemination, creating and spreading their own unique culture. Although the selfie has become a powerful tool for spreading youth subcultures and for the younger generation to find like-minded cultural arenas, it has even formed a technological barrier as a means of identity for the youth group.

Therefore, the natural pursuit of new technology by the new generation of the network provides an important basis for AR technology to intervene in the selfie field, and the technical upgrade of AR to the selfie design also provides core advantages to enhance the technology of the selfie and expand the dissemination channels of the selfie.

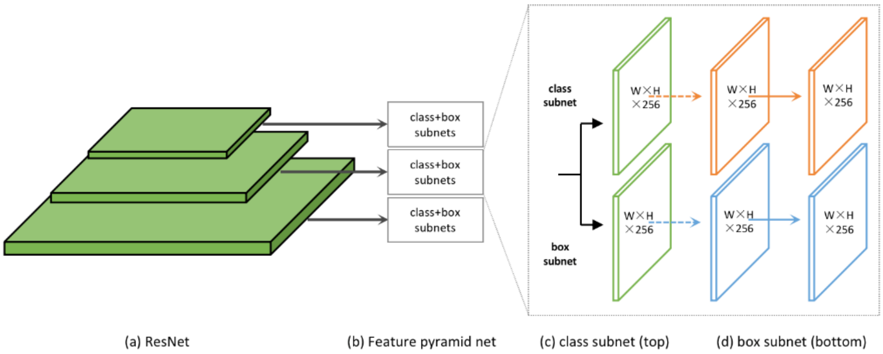
3. Methodology

In this paper, the RetainaNet model is selected as the base model for selfie design, which shows good performance in terms of speed and accuracy of object recognition.RetinaNet is a single-stage target detection algorithm, which consists of ResNet-N, Feature Pyramid Network (FPN), and two sub-networks of classification and regression.RetineNet in the training process Using Focal Loss to reduce the weights of simple samples and increase the weights of difficult samples so that the classifier can focus more on the classification of difficult samples, effectively solving the problem of positive and negative sample imbalance. The algorithm achieves faster detection speed and higher detection accuracy on PSACL and COCO datasets.

We introduces a self-attention mechanism module to improve the structure of the original model in order to further improve the detection accuracy. The self-attention mechanism in computer vision borrows from the selective attention mechanism in human vision, capturing the target regions in the image that need to be focused on, and then suppressing other useless information. The self-attentive mechanism calculates the response value of a location in a sequence by computing all locations and taking their weighted averages in the embedding space, which can be considered as a form of non-local mean. By introducing the self-attentive mechanism module, the network can enhance the learning ability among features, improve the discriminative ability between target and background classes, and improve the detection accuracy. In addition, this chapter uses the differential evolution algorithm (Berkemeier et al., 2019) to optimize the size of the Anchor, so that the Anchor with better size can match the corresponding feature map, and then improve the target recall rate.

3.1 RetinaNet Target Detection Model

The One-Stage target detection algorithm uses a priori frames to improve the detection speed, however, for a picture, thousands of candidate frames may be generated, and only a few of them contain valid targets, so if the candidate frames with targets are treated as positive samples and those without targets are treated as negative samples, it will lead to imbalance between positive and negative samples, and the detection accuracy will be low. In order to solve the problem of unbalanced positive and negative samples in the detection model, a new target detection scheme, RetinaNet, is proposed by Focal Loss, which can solve the problem of unbalanced positive and negative samples and greatly improve the detection accuracy by adjusting the weights of positive and negative samples and controlling the weights of easy to classify samples and hard to classify samples. The network structure of RetinaNet is shown in Figure 1. Its overall is composed of a backbone feature extraction network ResNet, a feature pyramid network FPN, and two sub regression networks.



(a) ResNet (b) Feature pyramid net (c) class subnet (top) (d) box subnet (bottom)

Figure 1. RetinaNet model

The experiments show that increasing the number of layers will lead to degradation of the network: for the traditional convolutional neural networks such as VGG-19 and GoogleNet-22, the loss of the model decreases gradually with the increase of the number of layers and is in a stable trend, but if the depth of the network continues to increase, the loss of the model will rebound, increase gradually and does not recover with the training time. The main feature extraction network of RetinaNet model, ResNet, is composed of residual blocks, which can deepen the convolutional neural network without network degradation.

The structure of Conv Block is shown in Figure 2, and its input dimension is different from the output dimension. The structure of the Identity Block is shown in Figure 2, and its input dimension is the same as the output dimension.

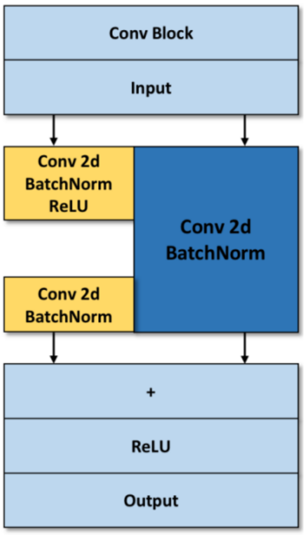


Figure 2. Conv Block Structure Schematic

3.2 Self-attentive Mechanism Module

Since the convolution operation is a local operation, the image features are extracted from the local area, and the information correlation between the local area and the whole is not available. Therefore, it is easy to be disturbed by the complex areas in the front background, which affects the detection accuracy. In order to solve this problem, we consider introducing a self-attentive mechanism to improve the feature extraction network. The self-attention model is based on the traditional visual attention model, which calculates the global pixel response to local pixels and can improve the global dependence learning ability among features [16]. The proposed self-attentive mechanism module is shown in Figure 3.

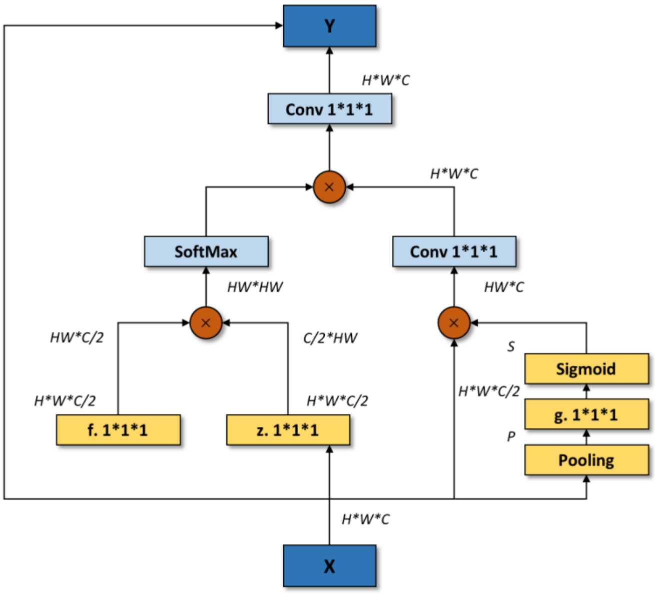


Figure 3. Self-attentive mechanism module

Figure 3 shows the improved self-attentive mechanism module in this paper, whose structure is similar to the residual block. where f(x), z(x), and g(x) are defined as 1×1 convolution kernels for cross-channel information fusion of Feature Map, which is computed by the following equation (1).

$$f(x) = W_f(x), z(x) = W_z(x), g(x) = W_g(x) \tag{1}$$

Where W_f, W_z and W_g are the weight parameters used to calculate the similarity of the corresponding feature points in the spatial location of the feature map, and the expression of similarity calculation is shown in equation (2).

$$S_{ij} = f(x_i)^T \varphi(x_j) \tag{2}$$

where i is used to represent the index values in the feature map, respectively, xi is used to represent the value of index i in the feature map, and s_{ij} is used to calculate the feature similarity between the position of j in the index and the position with the corresponding index in index i.

In this paper, we propose a non-local block that introduces the channel annotation mechanism into the non-local block. First, the global channel information is compressed into one channel by pooling the input X on average and then the channel representation P is mapped nonlinearly using the convolution of 1×1 and the Sigmoid activation function. The channel attention vector S is calculated as (3).

$$S = r(Wx) \tag{3}$$

where W is the weight of the 1×1 convolutional layers and r denotes the Sigmoid function.

The design of this paper follows the bottleneck structure, after obtaining the attention vector S of the channels, multiplying S with the original image X on each channel to obtain the feature map X1 , as shown in Equation (4).

$$K_c = S_c \cdot X_c \tag{4}$$

Based on the non-local operation in Non-local, the non-local operation to calculate the sum of feature weights at all locations in the feature map and the attention response for a given local region is shown in Equation (5).

$$Y_i = \frac{1}{C(x)} \sum_{vj} f(x_i, x_j) g(x_j) \tag{5}$$

where *x* is the input signal (RGB image); *Y* is the output signal of the same size as the x dimension. *f(x)* is used to calculate the feature similarity between a location *i* and all locations *j*. The function *g(x)* calculates the input signal representation at position *j*. Finally, the output signal *Y* is obtained by normalizing the response factor *C(x)*. The correlation matrix *M* between different positions is learned under supervision using the GroundTrue of the training sample, and is matrix multiplied with the feature map *K* after the channel calculation to obtain the output result of the self-attentive module. Based on the training sample GroundTrue, the corresponding attention can be calculated by scaling it to *H*×*W* . GroundTrue to determine whether the corresponding two pixel points belong to the same region of the important object, and the calculation formula is as in equation (6).

$$Y_i = \begin{cases} 0 & i, j \notin T \\ 1 & i, j \in T \end{cases} \tag{6}$$

In equation (6), the correlation is expressed as 1 if both pixels belong to the region of the important target, and 0 if both pixels do not belong to the important target region. The improved Non-Local Block improves the network's ability to use image channels and spatial information, and better guides the network to focus on the region of interest.

3.3 Confidence Interval

Intersection over Union (IoU) is the most commonly used evaluation metric in target detection benchmarks, and most of the current target detection algorithms use the intersection ratio of the prediction frame and the true frame to determine whether they fit or not (Kikuchi et al., 2022). However, the optimization degree of the two is not equal, and there is a case that the IoU of the prediction frame and the real frame are not equal when the l2-parameter of the two frames are equal. Therefore, the GIoU is used to measure the fit of the prediction frame to the true frame: for the prediction frame A and the true frame B, the minimum convex set C can be calculated, and the IoU and GIoU can be calculated as shown in equations (He, 2019) and (Grubert et al., 2017), respectively.

$$IoU = \left| \frac{A \cap B}{A \cup B} \right| \tag{7}$$

$$GIoU = IoU - \frac{|C/A \cup B|}{|C|} \tag{8}$$

GIoU can be expressed as equation (Mamone et al., 2020).

$$L_{GIoU} = 1 - GIoU \tag{9}$$

Therefore, from equation (Mamone et al., 2020), we can see that -1≤GIoU≤1. When the prediction frame A and the real frame B are equal, we have GIoU=1. When IoU = 0, i.e., there is no overlap between the prediction frame A and the real frame B, the GIoU formula can be transformed into equation (Yang et al., 2022).

$$GIoU = IoU - \frac{A \cup B}{C} - 1 \tag{10}$$

The closer the prediction frame is to the real frame, the closer the value of GIoU is to 0, while A U B remains unchanged, so we can keep optimizing to make C smaller, so that the prediction frame and the real frame become closer and closer.

3.4 Focal Loss Loss Function

In the One-Stage target detection algorithm, the significant difference in the number of positive and negative samples is one of the main factors leading to the low accuracy (Monterubbianesi et al., 2022). In the SSD target detection framework, the ratio of positive and negative samples is set to 3:1 in order to adjust the ratio. The Focal loss is actually a further improvement of the standard cross-entropy loss function, which is calculated as in equation (Cao & Cerfolio, 2019).

$$\mathrm{FL}(p_t) = -a_t(1 - p_t)^y log(p_t) \tag{11}$$

In equation (Cao & Cerfolio, 2019): a_t is the weight parameter between categories (0-1 dichotomous); $(1-p_t)^y$ is the simple/difficult sample adjustment factor; -log(p_t) is the initial cross-entropy loss function; *y* is the focusing parameter (Focusing Parameter).

3.5 Anchor Improvements

Anchor Boxes are used to predict Bounding Boxes. 128×128, 256×256, 512×512 are set in the Faster R-CNN target detection frame, and the three scale transformations 1:1, 1:2, 2:1 are combined to generate a total of 9 Anchor Boxes. In the Faster R-CNN target detection framework, three sizes of 128×128, 256×256, 512×512 are set, and a total of 9 Anchors are combined with three scale transformations 1:1, 1:2, 2:1 to predict the boxes. The Anchor mechanism is also used in YOLOv2, and the clustering algorithm K-means is used to optimize the Anchor size, and the experiments show that the Anchor Boxes obtained by K-means are more accurate than the hand-selected a priori boxes. Therefore, the selection of the optimal Anchor is also an important part to improve the accuracy.

In the experiments, it is found that after the introduction of the self-attentive mechanism, each point has to capture the global information relative to the whole feature map, which leads to a large amount of computation and memory capacity of the self-attentive module, so we consider optimizing the Anchor to reduce the computation and memory consumption. In this paper, we use the difference evolution algorithm (DE) (Sung et al., 2022) to optimize the ratio and scale of the Anchor on the validation set. We take the average of 9 different sizes from all the bounding boxes in the dataset, and then use these 9 bounding boxes as Anchor, so that it is faster and more accurate to fit the Bounding box of the detection target (van Nuenen & Scarles, 2021).

It was found that the default Anchor size = [32, 64, 128, 256, 512], Strides = [8, 16, 32, 64, 128], Aspect ratio (1:2, 1:1, 2:1) and Scales (20/3, 21/3, 22/3) did not produce the optimal accuracy. In this paper, a suitable Anchor is selected for training among all the labeled Bounding Boxes in the PASCAL dataset by using the DE algorithm. The algorithm iteratively improves the overall candidate for the objective function, and the goal is to select the best anchor for 3 sets of Scales The objective is to select the best Anchor settings with 3 sets of Scales and 3 sets of Aspect ratios, maximizing the number of Anchors between the target Bounding Box and the validation set. The objective is to select the best Anchor settings for 3 sets of Scales and 3 sets of Aspect ratio, maximizing the overlap between the target bounding box and the best Anchor Box on the validation set (Figure 4).

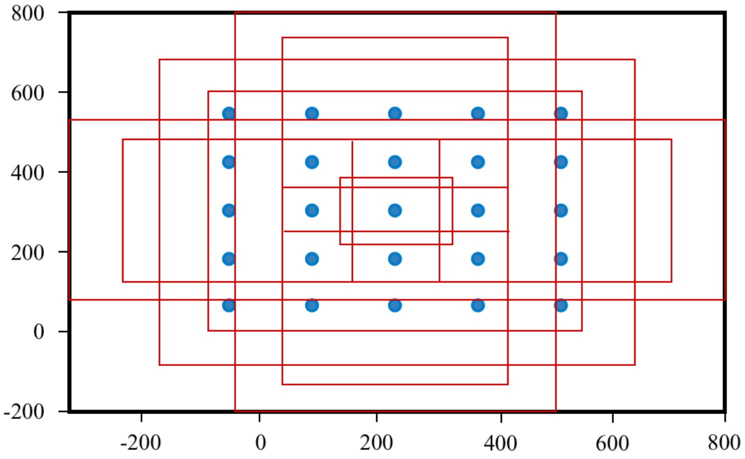


Figure 4. Improved anchor

3.6 Improved RetinaNet Model

In this paper, we use ResNet-50 as the backbone network, where C1, C2, C3, C4 and C5 are the feature layers obtained from ResNet-50 at different scales. The size of the input image of the model is set to 600 * 600 * 3 to reduce the computational effort. Only C3, C4 and C5 feature layers are passed to SAM in the experiment. SAM is used to understand the correlation between image space and channel information. Then, SAM_1, SAM_2 and SAM_3 are passed to FPN for multi-scale fusion, and P3, P4 and P5 are passed to Class network and Box network for convolution, which are used as the prediction results of classification regression and location regression, respectively.

In this paper, the SAM-RetinaNet model is constructed by combining the self-consciousness mechanism model with the ResNet-50 network model. The improved model is shown in Figure 5. The main body of the figure is constructed by the backbone feature extraction network ResNet-50, the feature pyramid network FPN, and two sub-networks of classification and regression. The feature pyramid fuses the feature layers output from the basic backbone feature extraction network in a top-up, front-to-back structure, so that each layer of the feature pyramid can be used to detect objects of different scales.

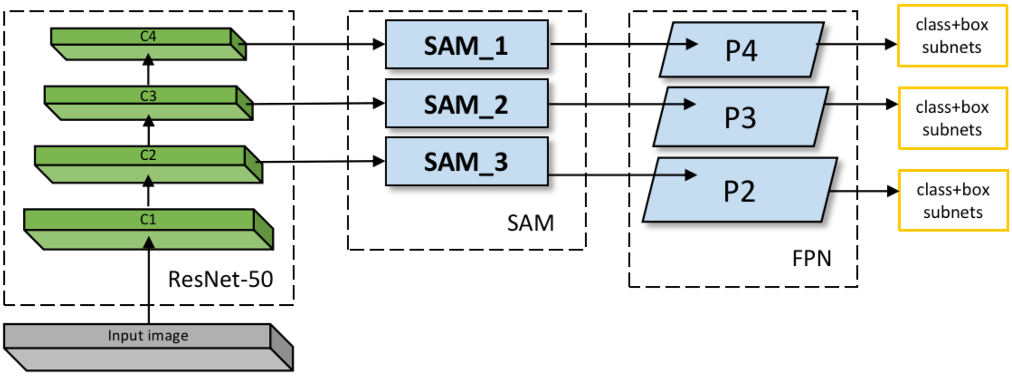


Figure 5. Improved RetinaNet model

4. Discussion and Results

In this paper, the model is trained on the EC1, EC2 and EC3 training sets based on the RetinaNet improved network architecture, and the experiments are conducted on the basis of the pre-trained weight files of the ResNet-50 network for faster convergence. This experiment uses the stochastic gradient descent method for training, and the learning rate is 1e-5. The patience is set to 2, i.e., when val_loss does not decrease for two epochs, the learning_rate is changed to half of the original one; when val_loss does not change for 10 epochs, the training is finished.

In this paper, the improved RetinaNet model is trained on the EC1 training set and EC2 and EC3 training set, and the total error of the model is stabilized around 0.9. In the figure, ResNet-50 represents the original ResNet-50+FPN network model, ResNet50-SA represents the network model with the Self-Attention layer added to the structure of ResNet-50+FPN, and ResNet50-SA_config represents the network model with the DE algorithm on top of ResNet-50+FPN. algorithm to optimize the Anchor size. The Classification Loss, Regressision loss, Loss variation curves and mAP curves of the original model and the improved two models are shown in Figure 6.

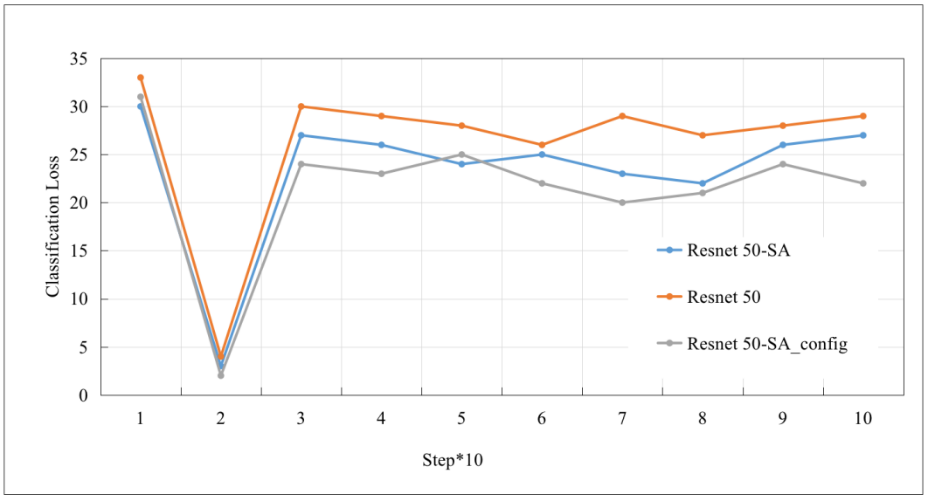


Figure 6. Comparison of regression loss curve

The experimental results show that the proposed model has significantly improved the target detection accuracy on the PASCAL dataset by introducing the self-attentive mechanism module and optimizing the Anchor with the differential evolution algorithm. The SAM-RetinaNet model with the optimized Anchor configuration on the basis of the attention mechanism achieves 87% mAP in the PASCAL dataset, and the experimental results demonstrate that the improved network model in this paper can effectively improve the target detection accuracy.

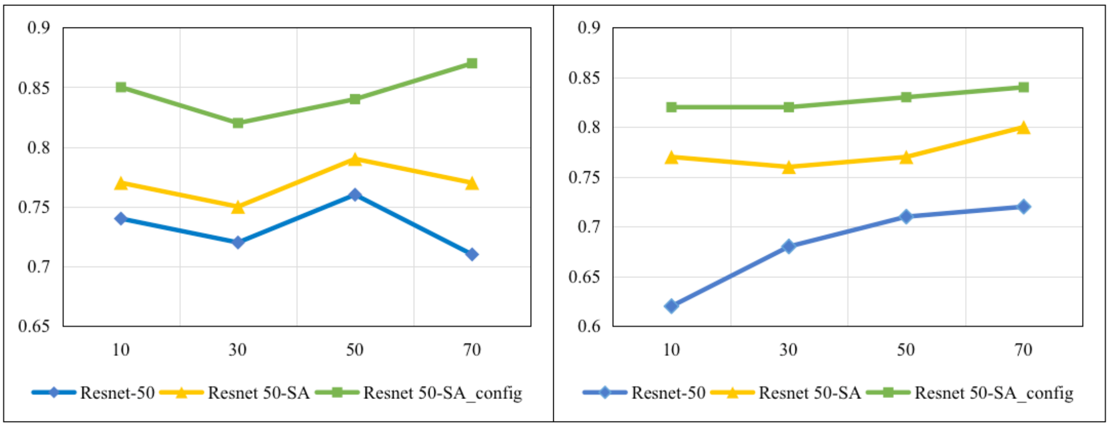


Figure 7. Comparison of experimental results

As shown in Figure 7, the detection accuracy of the improved RetinaNet model can reach 89.3% on the EC1 test set, which is 7.1 percentage points better than the original RetinaNet; and the detection accuracy on the EC2 and EC3 test set can reach 87.4%, which is 7.5 percentage points better than the original RetinaNet.

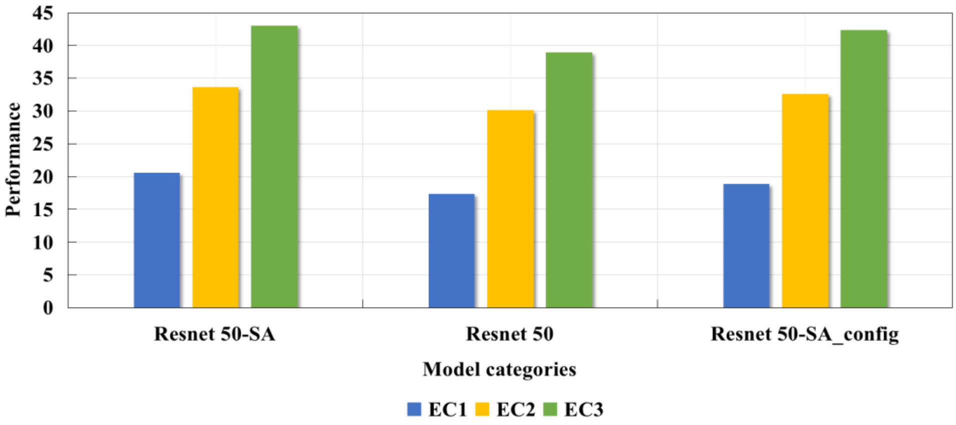


Figure 8. Comparison of experimental results

The improved model based on RetinaNet was trained on EC1, EC2 and EC3 sets, and the experimental results were compared with other common models as shown in Figure 8.

5. Conclusion

In AR task systems, the core problem is the alignment of virtual information and real scenes, and the accurate estimation of real targets is the key. Traditional target detection requires manual extraction of target features, which leads to high time complexity and low robustness of such algorithms, and the detection accuracy is greatly affected by lighting conditions and occlusion factors, which poses a challenge to how AR can be applied to selfies. Deep learning-based target detection technology can effectively solve the defects of traditional detection algorithms, and this paper uses convolutional neural networks to replace the process of manually designing features and extracting features, which can improve the target detection accuracy and real-time performance. Therefore, this paper deeply investigates the target detection algorithm based on deep learning and applies deep learning to selfie design, providing a specific technical path for the integration of AR and selfie. This paper proposed an improved RetinaNet model. The model's detection accuracy and robustness are effectively improved by introducing a self-attentive mechanism model to enhance the extraction and learning ability of image features, and the Anchor size in the dataset is optimized by using the differential evolution algorithm to ensure high detection efficiency while improving accuracy. In this paper, the improved model is compared with the PASCAL dataset, and the experiments show that the proposed model has better robustness in the presence of target occlusion and complex backgrounds in the images. In conclusion, the proposed model provides an important basis for the design optimization and future tradition of self-timer.

Conflict of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The data used to support the findings of this study are available from the corresponding author upon request.

Funding

This work did not receive any funding.

References

Aydoğdu, F. (2021). Augmented reality for preschool children: An experience with educational contents. *British Journal of Educational Technology*, *53*(2), 326-348. https://doi.org/10.1111/bjet.13168

Baran, B., Yecan, E., Kaptan, B., & Paşayiğit, O. (2019). Using augmented reality to teach fifth grade students about electrical circuits. *Education and Information Technologies*, *25*(2), 1371-1385. https://doi.org/10.1007/s10639-019-10001-9

Berkemeier, L., Zobel, B., Werning, S., Ickerott, I., & Thomas, O. (2019). Engineering of augmented reality-based Information Systems. *Business &amp; Information Systems Engineering*, *61*(1), 67-89. https://doi.org/10.1007/s12599-019-00575-6

Braly, A. M., Nuernberger, B., & Kim, S. Y. (2019). Augmented reality improves procedural work on an International Space Station Science Instrument. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(6), 866-878. https://doi.org/10.1177/0018720818824464

Cao, C., & Cerfolio, R. J. (2019). Virtual or augmented reality to enhance surgical education and surgical planning. *Thoracic Surgery Clinics*, *29*(3), 329-337. https://doi.org/10.1016/j.thorsurg.2019.03.010

Grubert, J., Langlotz, T., Zollmann, S., & Regenbrecht, H. (2017). Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *23*(6), 1706-1724. https://doi.org/10.1109/tvcg.2016.2543720

He, T. (2019). The sentimental fools and the fictitious authors: Rethinking the copyright issues of AI-generated contents in China. *Asia Pacific Law Review*, *27*(2), 218-238. https://doi.org/10.1080/10192557.2019.1703520

Kikuchi, N., Fukuda, T., & Yabuki, N. (2022). Future landscape visualization using a city digital twin: Integration of Augmented Reality and drones with implementation of 3D model-based occlusion handling. *Journal of Computational Design and Engineering*, *9*(2), 837-856. https://doi.org/10.1093/jcde/qwac032

Liu, J., & Keane, H. (2020). Naked loan selfies: Becoming collateral, becoming pornography. *New Media &amp; Society*, *23*(12), 3616-3633. https://doi.org/10.1177/1461444820957257

Lu, F., & Chia, S. C. (2022). When virtual makeovers become “real”: How SNS interactions drive selfie editing and Cosmetic Surgery. *Chinese Journal of Communication*, *16*(1), 73-89. https://doi.org/10.1080/17544750.2022.2085127

Lyu, Z., Jiao, Y., Zheng, P., & Zhong, J. (2021). Why do selfies increase young women’s willingness to consider cosmetic surgery in China? the mediating roles of body surveillance and body shame. *Journal of Health Psychology*, *27*(5), 1205-1217. https://doi.org/10.1177/1359105321990802

Mamone, V., Ferrari, V., Condino, S., & Cutolo, F. (2020). Projected augmented reality to drive osteotomy surgery: Implementation and comparison with video see-through technology. *IEEE Access*, *8*, 169024–169035. https://doi.org/10.1109/access.2020.3021940

Monterubbianesi, R., Tosco, V., Vitiello, F., Orilisi, G., Fraccastoro, F., Putignano, A., & Orsini, G. (2022). Augmented, virtual and mixed reality in Dentistry: A narrative review on the existing platforms and future challenges. *Applied Sciences*, *12*(2), 877. https://doi.org/10.3390/app12020877

Niu, G., Sun, L., Liu, Q., Chai, H., Sun, X., & Zhou, Z. (2019). Selfie-posting and young adult women’s restrained eating: The role of commentary on appearance and self-objectification. *Sex Roles*, *82*(3-4), 232-240. https://doi.org/10.1007/s11199-019-01045-9

Sung, E. (Christine), Danny Han, D.-I., Bae, S., & Kwon, O. (2022). What drives technology-enhanced storytelling immersion? the role of Digital humans. *Computers in Human Behavior*, *132*, 107246. https://doi.org/10.1016/j.chb.2022.107246 van Nuenen, T., & Scarles, C. (2021). Advancements in technology and digital media in Tourism. *Tourist Studies*, *21*(1), 119-132. https://doi.org/10.1177/1468797621990410

Wu, M. S., Song, C., & Ma, Y. (2019). Selfie taking may be nonharmful: Evidence from adaptive and maladaptive narcissism among chinese young adults. *Human Behavior and Emerging Technologies*, *1*(3), 240-244. https://doi.org/10.1002/hbe2.166

Yang, M., Wang, S., Zhang, N., Zhou, A., & Ma, X. (2022). Survey on tracking and registration technology for Mobile Augmented Reality. *International Journal of Web and Grid Services*, *18*(2), 99. https://doi.org/10.1504/ijwgs.2022.121960

Zheng, D., Ni, X., & Luo, Y. (2018). Selfie posting on social networking sites and female adolescents' self-objectification: The moderating role of Imaginary Audience Ideation. *Sex Roles*, *80*(5-6), 325-331. https://doi.org/10.1007/s11199-018-0937-1

Copyrights