Improved image semantic segmentation with domain adaptation for mechanical parts

Xin XIE*, Yuhui HUANG*, Tiancheng WAN*, Lei XU* and Fengping HU**

* School of Information Engineering, East China Jiaotong University
808 Shuanggang East Street, Economic and Technological Development Zone, Nanchang 330013, China
E-mail: 2857060768@qq.com

** School of Civil Engineering and Architecture, East China Jiaotong University
808 Shuanggang East Street, Economic and Technological Development Zone, Nanchang 330013, China

Abstract
Non-contact detection methods based on computer vision are widely used in industrial production. When collecting images, different light source schemes are often used to meet the detection requirements of different objects. The image styles collected under different lighting scenes are diverse, and the image semantic segmentation model trained with a specific dataset has unsatisfactory performance when processing images in different domain. This paper designs a mechanical part image semantic segmentation model based on domain adaptation and GAN. In order to verify the effectiveness of the proposed method, this paper collects and labels some gear images and constructs a dataset. The first step is to train encode-decoder structure with an enhanced memory module with source dataset to achieve semantic segmentation, then the dataset is transformed into a target domain by GAN, and finally the image semantic segmentation model is fine-tuned with the target dataset. The advantage is that since the cycle-consistency loss is used to constrain the spatial structure of the reconstructed image, the two datasets can share a semantic segmentation label. Experiments show that the image semantic segmentation method based on domain adaptation has achieved good results on different styles of gear parts image datasets, and achieved a pixel accuracy of 94.1%.

Keywords: Domain adaptation, Memory module, Encode-decoder network, Semantic segmentation, Generative adversarial networks

1. Introduction

With the development of manufacturing technology, the basic components of sophisticated equipment, such as slender shafts, gears, etc., are used more and more widely (Chen and Shao, 2011). In the industrial production process, affected by the processing conditions of equipment, cutting vibration and thermal deformation and other factors, the machining accuracy of parts often can’t fully meet the requirements, and the accuracy of part dimensions directly affects the efficiency and life of the equipment, so the factory inspection of parts is a very important link (Zhi, 2019). Because the non-contact detection method has the advantages of non-destructive, high-efficiency, unified standards, etc., the detection method based on computer vision is widely used in the field of factory inspection in the industry. This method usually first extracts the contour edge of the target to be detected in the image, and then calculates the actual size according to the space-pixel conversion relationship obtained by the camera calibration (Xie et al., 2019a). Therefore, the quality of the edge of the extracted target object directly affects the result of size detection. According to the realization principle of edge detection technology, it can be divided into traditional edge detection methods and deep learning methods. Traditional edge detection methods include: differential operator algorithm (Tang et al., 2011), mathematical morphology method (Zhao et al., 2020), bionic algorithm (Xie et al., 2019b) and so on. These methods mostly use the low-level visual information of the image, such as color, shape, texture and roughness information for processing, and have certain effects for specific scenes, but they are still difficult to meet the needs in the face of
complex scenes. The edge detection method based on deep learning is realized by image semantic segmentation, that is, the image is divided into two or more semantically related regions. Among them, semantic segmentation methods based on pixel classification include: FCN method (Long et al., 2015), encoder-decoder method (Badrinarayanan et al., 2017), feature fusion method (Dong et al., 2021), GAN method (Reza et al., 2019) and so on. Compared with traditional methods, deep learning method has a good effect in detecting complex image edges, and lots of the work are carried on how to improve the segmentation quality (Ghosh et al., 2019).

Deep learning is a data-driven method, which requires a large amount of data and corresponding labels to train the neural network, and it is an extremely time-consuming process to make real labels for these data (Soares et al., 2019). In the field of industrial production, different light source schemes will be arranged for different detection targets, such as light frequency and color temperature, so the image styles of mechanical parts collected are different. With many existing semantic segmentation methods (Wang et al., 2019, Ronneberger et al., 2015, Vuola et al., 2019), semantic segmentation can achieve a high accuracy by training the model in a supervised way. However, when the trained model deals with new datasets, the index of semantic segmentation decreases dramatically (Gong et al., 2021). Therefore, how to improve the generalization ability of neural networks is a great challenge, and unsupervised methods are worth exploring.

Generative Adversarial Networks (GAN) is an unsupervised deep learning method. GAN-based image segmentation algorithms are rarely studied in the field of industrial production. The generalization ability of existing models is not satisfactory, and it is difficult to meet different detection requirements. This paper focuses on semantic segmentation of mechanical parts image. Since a large number of supervised learning samples are not available in many applications, GAN is used for domain adaptation after labelling a small number of sample images to improve the generalization ability of the model. In this paper, we train a semantic segmentation network with encoder-decoder structure. Firstly, the feature decoding process is reasonably restricted by the memory module, and then the source image dataset is reconstructed to the target domain using GAN, which the spatial structure of the image is preserved. Finally, fine-tune the semantic segmentation model with the target dataset and source domain labels to complete the semantic segmentation task of images in the target domain.

The contribution of this paper are as follows:
1) Design a semantic segmentation model enhanced by memory module;
2) Collected and labeled some gear images to build a dataset;
3) The effectiveness of the proposed method in image processing is proved, and the semantic segmentation model can be adapted to different types of industrial parts images.

The rest of the paper is organized as follows: The related work of domain adaptation, memory module and GAN are introduced in Section 2. The GAN improved by memory module in Section 3. The semantic image segmentation method and model optimization process of domain adaptation Introduced in Section 4. The experiments and results are discussed in Section 5. Finally, the conclusion of this paper is presented in Section 6.

2. Related work

Domain adaptation is a representative method in transfer learning, which refers to the use of information-rich source domain samples to improve the performance of the target domain model. The source domain represents a domain different from the test sample, but has rich supervision information; the target domain represents the domain where the test sample is located, with no labels or only a few labels. The source domain and the target domain often belong to the same type of task, but the distribution is different. Concerning the domain adaptation for semantic segmentation, many works on this field focused on simulated data (Vazquez et al., 2014, Peng et al., 2018). Hoffman et al. (Hoffman et al., 2017) proposed a novel discriminatively-trained Cycle-Consistent Adversarial Domain Adaptation model. Multiple loss functions are used by the model to train the model, and the feature level and pixel level are aligned at the same time to improve its performance on the target images. Tsai et al. (Tsai et al., 2018) proposed a domain adaptation method for pixel-level semantic segmentation via adversarial learning. They found that this approach improves the performance of the model on target domains. Zhang et al. (Zhang et al., 2018) proposed a novel deep learning model, Fully Convolutional Adaptation Networks (FCAN), which is composed of Appearance Adaptation Networks (AAN) and Representation Adaptation Networks (RAN). This method improved the adaptation performance of the network. Most of the existing domain adaptation methods are used in the fields of medicine and remote sensing.
images. The paper innovatively applies the domain adaptation method to industrial parts images, and has achieved excellent results compared with other methods.

Traditional deep learning models use the attention mechanism as their memory function, but the memory generated by this method is too small, and a lot of information is lost when the input information is encoded into a feature vector. Initially Weston et al. (Weston et al., 2015) proposed a memory module that can be flexibly operated. Then, memory modules have aroused widespread interest in solving different problems. Graves et al. (Graves et al., 2014) used context-based addressing and location-based addressing methods for external memory module addressing to enhance the memory ability of neural networks. Santoro et al. (Santoro et al., 2016) proposed a memory-enhancing neural network to quickly absorb the information contained in samples and use this information to make accurate predictions in situations where only a few samples are provided. Through the memory module, the paper can guide the output of the generator and segmentation network to make it close to the expected result.

Generative Adversarial Networks (GAN) is an unsupervised deep learning method that uses game theory to optimize model parameters through adversarial training (Goodfellow et al., 2014). Luc et al. (Luc et al., 2016) used the generator to predict the segmented image, and then compared the difference between the segmented image and the ground truth image through the discriminator. Using GAN for semantic segmentation can improve the high-order consistency between the label and the predicted value, and obtaining continuous segmentation result. Zaheer et al. (Zaheer et al., 2020) used inadequately trained generators and artificially forged abnormal images as negative samples to train the discriminator, and converted the target of the discriminator to judge the quality of the reconstructed images, which improved the stability of the model. Mirza et al. (Mirza et al., 2014) designed a Conditional GAN, in which the model generator and discriminator have additional constraint conditions as input to make the output of the model controllable. The image output by GAN is random, and the cycle-consistency loss function is used to obtain a new style image without changing the image structure.

3. Generative adversarial network improved by memory module

3.1 Memory module

Image semantic segmentation assigns a label to each pixel of the input image, and uses different colors to represent different types of pixels in the output of prediction image, so it can be regarded as a reconstruction of the input image. Reconstruction refers to encoding the image \( I \) to obtain the deep features, and then decoding the deep features to obtain the output image \( \hat{I} \). Generally, the reconstruction loss between \( I \) and \( \hat{I} \) is compared as the detection index (Ryan et al., 2019). In semantic segmentation, if the reconstruction loss between the reconstructed image \( \hat{I} \) and the ground truth label is smaller, the segmentation performance is better. For the image of mechanical parts, no matter how the style domain of the collected image is transformed, the semantic segmentation label can be represented by a binary image, that is, the style of the label image belongs to only one domain. When the image semantic segmentation model is trained with a dataset in specific domain, the model tends to fit the data distribution of the image in this domain in order to achieve the best segmentation performance, then the generalization ability is poor. The memory module (Dong et al., 2015) can reasonably restrict the feature decoding process and guide the model output to the data distribution of the label image, which can enhance the generalization ability of the model. Therefore, based on U-Net (Ronneberger et al., 2015), this paper adds a memory module to constrain the feature mapping of the semantic segmentation model: after the deep features are extracted from the encoding part, these features are mapped to the data distribution of semantic segmentation labels through the memory module, and then used as the input of the decoding part to predict the segmentation image. The overall structure is shown in Fig. 1.

The input of the U-Net model encoder is a single-channel image with a size of 512×512, followed by two convolution operations to obtain a 64-channel feature map with a size of 512×512. After the feature map is subjected to a 2×2 max pooling operation, its length and width are halved and used as the input of the next convolutional layer. After that, the above two convolutions and pooling are performed, which is repeated four times, and finally the feature \( x \) is obtained. In the U-Net model with memory module we proposed, the feature \( x \) obtained by the encoder is input to the memory module, and the obtained features can reasonably restrict the feature decoding process, which is beneficial to improve the segmentation. The feature \( \hat{x} \) output by the memory module is input to the decoder, and the resolution of the feature is gradually expanded through the upsampling operation, therefore the model outputs a 512×512 semantic segmentation image.
The memory module $M \in \mathbb{R}^{N \times C}$, which is a matrix containing $N$ real-valued vectors on a fixed dimension $C$. Assuming that the dimension $C$ is the same as the input feature $x$, then $X = \mathbb{R}^{C}$. Let $m_i$ denote the $i$-th row of matrix $M (i \in [N])$, that is, $m_i$ denote a memory item. The memory module uses the soft addressing vector shown in Eq. (1) to map the feature $x$ to $\hat{x}$:

$$\hat{x} = \alpha M = \sum_{i=1}^{N} \alpha_i m_i$$  \hspace{1cm} (1)

The weight vector $\alpha$ is calculated from the input feature $x$. It is a column vector with the sum of non-negative items equal to 1, and $\alpha_i$ represents the $i$-th item of $\alpha$. Equation (2) shows that access to memory module $M$ requires addressing weight $\alpha$. The memory module $M$ is used to record the normal mode of data during training, and $\alpha_i$ can be calculated according to the similarity between the memory item and the code $x$:

$$\alpha_i = \frac{\exp(d(x, m_i))}{\sum_{j=1}^{N} \exp(d(x, m_j))}$$ \hspace{1cm} (2)

Where $d(\cdot, \cdot)$ is the similarity measure, calculated by cosine similarity:

$$d(x, m_i) = \cos(x, m_i) = \frac{x \cdot m_i^T}{\|x\|_2 \|m_i\|_2}$$ \hspace{1cm} (3)

The memory module maps the coded $x$ to $\hat{x}$ according to the Eq. (1), memory module limits each mapping can only access the specified memory items. When training the model, the decoding module can only use a small number of memory items to reconstruct the data, forcing the memory module to record the most representative original data feature distribution. The parameters of the memory module will no longer be updated in the test phase, so as to constrain the reconstruction of the test data to the feature distribution direction of the training data.

Since the weight vector $\alpha$ is densely distributed, some abnormal data can be reconstructed into normal data through a complex combination of memory items. Therefore, hard shrinkage is used to enhance the sparsity of the weight vector $\alpha$.

---

**Fig. 1 Improved semantic segmentation model.**
\[
\hat{\alpha}_i = h(\alpha_i, \lambda) = \begin{cases} 
\alpha_i, & \alpha_i > \lambda \\
0, & \text{other} 
\end{cases} \quad \lambda \in \left[\frac{1}{N}, \frac{3}{N}\right]
\]  

(4)

\(\hat{\alpha}_i\) is the i-th item of memory module addressing weight after shrinking, and \(\lambda\) is the vector of shrinkage threshold. In order to be trained in the neural network model, Eq. (4) is mapped to a continuous function:

\[
\hat{\alpha}_i = \frac{\max (\alpha_i - \lambda, 0) \cdot \alpha_i}{|\alpha_i - \lambda| + \varepsilon}
\]  

(5)

\(\varepsilon\) is a very small positive number set to prevent division by zero. After the shrink operation, normalize \(\alpha_i\) to facilitate feature mapping:

\[
\hat{\alpha}_i = \frac{\hat{\alpha}_i}{\|\hat{\alpha}\|}, \quad \forall i
\]  

(6)

The normalized weight vector \(\alpha\) was substituted into Eq. (1) to complete the feature mapping task. When training the semantic segmentation model, suppose the dataset contains M images, the label image corresponding to each segmentation prediction map \(x_m\) is \(\hat{x}_m\), and the segmentation error is measured by the L2 norm:

\[
D(x_m, \hat{x}_m) = \|x_m - \hat{x}_m\|^2_2
\]  

(7)

Assume that \(\hat{\alpha}_m\) is the addressing weight of each sample image \(x_m\). Considering that all terms of \(\hat{\alpha}\) are non-negative and \(\|\hat{\alpha}\|=1\), minimizing the entropy of \(\hat{\alpha}_m\) further improves sparsity:

\[
E(\hat{\alpha}_m) = \sum_{i=1}^{I} -a_i \cdot \log(a_i)
\]  

(8)

Where \(I\) is the dimension of the weight vector, \(a_i\) is the value of the i-th dimension in the weight vector \(\hat{\alpha}_m\). The training of the semantic segmentation model aims at minimizing the Eq. (9). \(\gamma\) is a hyper-parameter in training. In experiment, \(\gamma=0.0002\) leads to desirable results in all our experiments.

\[
\text{Loss} = \frac{1}{M} \sum_{m=1}^{M} (D(x_m, \hat{x}_m) + \gamma E(\hat{\alpha}_m))
\]  

(9)

### 3.2 Generative adversarial network

This paper designs a Generative Adversarial Network, which aims to simulate the images of mechanical parts collected under different light sources without pairing images of different styles. The generator \((G_{AB})\) reconstructs the image data from domain A to the domain B, then the discriminator \((D_B)\) compares the difference between the generated image B and the real image B, so as to optimize the generator performance. The model trained with cycle-consistency loss (JY et al., 2017) to constrain the reconstructed image, constrain the pixel coordinates and range of the part target in the output image, which makes it consistent with the input image and ensures the reliability of size detection after the part image style is transferred.

The generator of the model use the U-Net enhanced by the memory module in Section 2, and the performance of the generator can be improved by changing the data distribution parameters of the memory module. In the discriminator part, the model adopts the structure of the encoder. After the input image is encoded into a one-dimensional vector, the Sigmoid activation function is used to obtain the binary output to identify the authenticity. The model structure is shown in Fig. 2.

Suppose \(\{a_i\}_{i=1}^{N} \in A\), \(p_{data}(a)\) is the data distribution of image a, the data distribution is \(a \sim p_{data}(a)\);
similarly, there are \( \{b_i\}_{i=1}^N \in B \) and the data distribution \( b \sim p_{\text{data}}(b) \). The model optimizes the generator by adversarial loss. The goal for generator \( G_{AB} \) is to minimize Eq. (10), while discriminator \( D_b \) maximizes Eq. (10).

\[
L_{AB} (G_{AB}, D_b, A, B) = E_{b \sim p_{\text{data}}(b)} \left[ \log D_b (b) \right] + E_{a \sim p_{\text{adv}}(a)} \left[ 1 - \log D_b (G_{AB} (a)) \right]
\]  

(10)

The definition form of loss \( L_{BA} \) of generator \( G_{BA} \) and discriminator \( D_A \) is similar to that shown in Eq. (10). The two parts can be combined and expressed as Eq. (11).

\[
\begin{align*}
&\min_{G_{AB}, \max_{D_b}} L_{AB} (G_{AB}, D_B, A, B) \\
&\min_{G_{BA}, \max_{D_A}} L_{BA} (G_{BA}, D_A, B, A)
\end{align*}
\]  

(11)

The task of data domain transfer can be completed by adversarial loss, but the target subject of the input image and the output image can’t be guaranteed to be consistent. The generators \( G_{AB} \) and \( G_{BA} \) are random functions, and the neural network can map the input image to any image in the target domain, and the ideal mapping relationship is Eq. (12):

\[
\begin{align*}
&a \rightarrow G_{AB} (a) \rightarrow G_{BA} (G_{AB} (a)) \approx a \\
&b \rightarrow G_{BA} (b) \rightarrow G_{AB} (G_{BA} (b)) \approx b
\end{align*}
\]  

(12)

For each image from data A, after one conversion period, the image should be restored to the original appearance, and the cycle-consistency loss can be used to constrain the above process:

\[
L(G_{AB}, G_{BA}) = E_{a \sim p_{\text{adv}}(a)} \left[ \|G_{BA} (G_{AB} (a)) - a\| \right] + E_{b \sim p_{\text{data}}(b)} \left[ \|G_{AB} (G_{BA} (b)) - b\| \right]
\]  

(13)

For the entire model, the goal is to optimize the loss function Eq. (14).

\[
\text{Loss} = L_{AB} (G_{AB}, D_B, A, B) + L_{BA} (G_{BA}, D_A, B, A) + \gamma \cdot L(G_{AB}, G_{BA})
\]  

(14)

where \( \gamma \) is the weight of the cycle-consistency loss, which determines the similarity of the target structure after the data domain translated by GAN.
4. Method

4.1 Image semantic segmentation with domain adaptation

This paper designs a semantic segmentation model of part image with domain adaptation based on GAN, as shown in Fig. 3.

Fig. 3 GAN-based part image domain adaptive semantic segmentation model. The process of image processing can be divided into four steps: 1) The source image datasets are used to train the memory module enhanced semantic segmentation model based on U-Net; 2) GAN converts the source image datasets to the target domain, only changing the image style and preserving the image spatial structure; 3) Use the target domain image generated by GAN to fine-tune the semantic segmentation model in step 1 to fit the data distribution of the target domain; 4) Use the fine-tuned enhanced U-Net model to process the mechanical part images in the new target domain datasets.

4.2 Optimization of model

Suppose the source domain image is \(X_A\), the target domain image is \(X_B\). \(X_B\) is generated by \(X_A\), with the same spatial structure and different style, so they can share a semantic segmentation label \(Y_A\).

4.2.1 Training the semantic segmentation model

First use the source domain dataset \(X_A\) to train the semantic segmentation model \(M_A\), and use the cross-entropy to optimize the loss function as Eq. (15).

\[
L_{M_A}(M_A, X_A, Y_A) = -\mathbb{E}_{(x, y) \sim (X_A, Y_A)} \sum_{c=1}^{C} \eta_{c \sim 0, 1}[\text{log}(\text{Sigmoid}(M_A^c(x_A)))]
\]  

Among them, \(\mathbb{E}_{(x, y) \sim (X_A, Y_A)}\) is the mathematical expectation obtained from the data distribution of \(X_A\) and \(Y_A\), \(C\) is the number of target categories, and \(\eta_{c \sim 0, 1}\) is the corresponding loss between category \(c\) and other categories. The dataset in this paper has two categories, target and background, so the Sigmoid activation function is used for binary classification. If there are multiple categories, the activation function can be set as Softmax (Wei et al., 2019).

4.2.2 Training domain shift model

In order to improve the generalization ability of the semantic segmentation model, this paper uses dataset with different domain to train the image semantic segmentation model. However, it is very time-consuming to make labels. Different styles of images are generated through GAN, and the cycle-consistency loss is used to constrain the image structure, so that images in different domain share the same semantic label.

The goal of generator \(G_{AB}\) is to reconstruct image \(X_A\) to the target domain \(X_B\) as much as possible, while the goal of discriminator \(D_B\) is to identify the real image \(X_B\) and generated image \(X_b\) as much as possible, so the goal is to optimize the Eq. (16):
\[
L(G_{AB}, D_B, X_A, X_B) = E_{x_a \sim X_a} \left[ \log D_B (x_a) \right] + E_{x_b \sim X_b} \left[ \log \left( 1 - D_B \left( G_{AB} (x_a) \right) \right) \right]
\]

(16)

For generator \( G_B \) and discriminator \( D_A \), there are:

\[
L(G_{BA}, D_A, X_B, X_A) = E_{x_a \sim X_a} \left[ \log D_A (x_a) \right] + E_{x_b \sim X_b} \left[ \log \left( 1 - D_A \left( G_{BA} (x_b) \right) \right) \right]
\]

(17)

In order to ensure the consistency of the reconstructed image structure and content, the whole GAN uses a cycle-consistency loss function, which is defined as Eq. (18).

\[
Loss(G_{AB}, G_{BA}, X_A, X_B) = E_{x_a \sim X_a} \left[ \left\| G_{BA} \left( G_{AB} (x_a) \right) - x_a \right\|_2 \right] + E_{x_b \sim X_b} \left[ \left\| G_{AB} \left( G_{BA} (x_b) \right) - x_b \right\|_2 \right]
\]

(18)

Minimize the above loss function and complete the optimization task of the image domain shift model.

### 4.2.3 Fine-tuning the semantic segmentation model

For similar datasets, the features of image extracted by CNN are mostly general, and only the last few layers are related to specific tasks. Fine-tuning (Tajbakhsh et al., 2016) is to update the parameters of last few layers of the neural network by using existing model structure and pre-trained parameters, which can reduce training costs while also obtaining good generalization capabilities. Taking the segmentation model \( M_A \) trained by source domain image \( X_A \) as the benchmark, and using the target domain image \( X_B \) to fine-tune \( M_A \), then the new model \( M_B \) is suitable for the new target domain. The loss function is defined as Eq. (19).

\[
L_{M_A} \left( M_B, X_B, Y_A \right) = -E_{(x, y) \sim (X_B, Y_B)} \sum_{c=1}^{C} \eta_{c} \log \left( \text{Sigmoid} \left( M_B \left[ c \right] (x) \right) \right)
\]

(19)

In this case, \( X_B \) is the target domain image generated by GAN, and other definitions are consistent with Eq. (15). Finally, the model \( M_B \) can be used to process parts images of different styles collected under new lighting environment.

### 5. Results and discussion

#### 5.1 Dataset

The goal of this paper is to detect the images collected on the industrial production line, so this paper constructs an image dataset of gear parts images. First, 600 source images are collected, which are taken in a relatively perfect state. When the image is collected in an industrial field, the temperature of the sensor is too high, and the collected images contain a lot of noise. Then 5% Gaussian noise is added through Photoshop to obtain noisy images. The expanded dataset contains a total of 1200 images. In the semantic segmentation label, the black area is the target and the white area is the background. Some examples of the dataset are shown in Fig. 5 (source image and label).

#### 5.2 Evaluation criteria

Semantic segmentation uses the common indexes pixel accuracy (PA) and intersection over union (IOU) to evaluate segmentation performance. For the images generated by GAN, this paper uses structural similarity (SSIM) (Alain et al., 2010) to measure the similarity of the two images after reconstruction, which is defined as Eq. (20).

\[
SSIM_{(x, y)} = \frac{2\mu_x \mu_y + \varepsilon_1 \left( 2\sigma_{xy} + \varepsilon_2 \right)}{\left( \mu_x^2 + \mu_y^2 + \varepsilon_1 \right) \left( \sigma_x^2 + \sigma_y^2 + \varepsilon_2 \right)}
\]

(20)
Where \( \left( x, y \right) \) represents two images to be compared, \( \mu_x \) and \( \mu_y \) are the mean values of the images, \( \sigma_x \) and \( \sigma_y \) are the variances of the images, \( \sigma_{xy} \) is the covariance of the images \( x \) and \( y \), and \( \varepsilon \) is the constant set to prevent the error of division by zero. The value range of SSIM is \([0,1]\), and the larger the value, the higher the similarity of the image.

Peak signal to noise ratio (PSNR) (Alain et al., 2010) is also an objective evaluation standard of images, which is used to measure image quality. It is defined as Eq. (21).

\[
PSNR = 10 \cdot \log_{10} \left( \frac{MAX^2}{MSE} \right)
\]  

(21)

Where \( MAX_x \) is the maximum pixel value of image \( x \), and MSE is the mean square error between image \( x \) and noiseless image \( \tilde{x} \).

5.3 Effectiveness of memory module

This experiment compares the segmentation effect of the original U-Net model and the enhanced model (U-Net model with memory module) in the gear part image dataset. The Enhanced U-Net model with memory module is better than the U-Net model in detail segmentation, the edge segmentation effect is more accurate. The results are shown in Fig. 4.

5.4 Comparison of image domain transformation

Taking the mechanical part image dataset as input, in order to make the generated image and the input image have a similar spatial structure, the cycle-consistency loss function is used between the two generators (as shown in Fig. 2, \( G_{AB} \) and \( G_{BA} \)) and the two discriminators (\( D_A \) and \( D_B \)). The experimental results are shown in Fig. 5.

Among them, the source image and label are the existing dataset, the domain shift image is the image generated by
GAN, and the reconstructed source image is the image with the source data style generated by GAN through domain shift image. The spatial structure of the reconstructed source image is the same as the source image, which eliminates most of the noise interference, but the edge of the part is also accompanied by some illumination effects of domain shift. The objective evaluation indexes of the image domain shift experiment are shown in Table 2:

It can be seen that the domain shift image is closer to the source image in structure, while the SSIM is slightly reduced after the reconstructed source image is generated twice. However, the domain shift image contains more illumination factors, and the reconstructed source image highlights the part target and ignores some background factors, so the PSNR index of reconstructed source image is better than that of source image.

![Fig. 5 Domain shift image experiment.](image)

### 5.5 Effectiveness of fine tuning the image semantic segmentation model

This experiment compared the semantic segmentation effect of the enhanced model in Section 2 on the domain shift dataset, there are four groups of control experiments. Case1: The original U-Net model is trained and tested with domain shift dataset. Case2: The memory module enhanced U-Net model is trained and tested with domain shift dataset. Case3: Freeze the parameters of the enhanced U-Net model trained by the source dataset, and directly process the domain shift dataset. Case4: The enhanced U-Net model trained by the source dataset is fine-tuned with the domain shift dataset, and then the segmentation task is carried out on domain shift dataset. The experimental results are shown in Table 3.

<table>
<thead>
<tr>
<th>Case</th>
<th>PA</th>
<th>IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case1</td>
<td>0.893</td>
<td>0.744</td>
</tr>
<tr>
<td>Case2</td>
<td>0.927</td>
<td>0.785</td>
</tr>
<tr>
<td>Case3</td>
<td>0.824</td>
<td>0.703</td>
</tr>
<tr>
<td>Case4</td>
<td>0.941</td>
<td>0.817</td>
</tr>
</tbody>
</table>

![Table 3 Controlled experiments on domain shift datasets.](table)
The results of Case1 and Case2 show that a considerable semantic segmentation effect can be obtained by training the model with target dataset. The results of Case3 show that semantic segmentation performance is poor when target data is processed with the old model. The results of Case4 show that the fine-tuning model has the best effect on semantic segmentation.

The fine-tuning semantic segmentation model can not only adapt to target data domains, but also has better effect by combining the characteristics of two data domains compared with the training of a single dataset.

6. Conclusion

In industrial production, there are different detection requirements for the images of mechanical parts, and different light source schemes will be designed. Aiming at the problem of insufficient generalization ability of existing models for parts images collected under different conditions, this paper proposes an image segmentation method based on domain adaptation. Using the existing dataset to initially train the semantic segmentation model, and then let the target domain dataset generated by GAN to fine-tune the model to realize the semantic segmentation task of the new target image dataset. Since the cycle-consistency loss function is used to constrain the spatial structure of the reconstructed image, the two datasets can share a semantic segmentation label.

Experimental results show that the proposed method is superior to the existing models on the target domain datasets. In addition, each image of the dataset in this paper only contains one part target. In actual detection, multiple targets are often collected in an image. In the future, the research will be carried out in the field of multi-objective instance segmentation in order to meet the requirements of multi-objective detection in factories.

Acknowledgement

This research was funded by the National Natural Science Foundation of China, under Grant No.61762037, No.62162026 and No.61872141, and the Science and Technology Key Research and Development Program of Jiangxi Province, under Grant No.20202BBEL53004.

References

JY, Z., Park, T. and Isola, P., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,


Xie, X., Ge, S. and Xie, M., An Improved Industrial Sub-Pixel Edge Detection Algorithm Based on Coarse and Precise Location (2019a), pp.1–10, Journal of Ambient Intelligence and Humanized Computing.


Zhi, S., Research on visual measurement method of medium and small module gears and tooth pitch measurement technology (2019), Shenyang University of Technology (in Chinese).