A Mechanical Parts Image Segmentation Method Against Illumination for Industry

Xin Xie (✉ xiexin@ecjtu.edu.cn)

East China JiaoTong University <https://orcid.org/0000-0002-3649-5741>

Tiancheng Wan

East China JiaoTong University

Bin Wang

East China JiaoTong University

Lei Xu

East China JiaoTong University

Xinlei Li

East China JiaoTong University

Research

Keywords: Joint Learning, Semantic Segmentation, Asymmetric Convolution, BN Fusion, Edge Detection

Posted Date: December 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-119471/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

A Mechanical Parts Image Segmentation Method Against Illumination for Industry

Xin Xie*, Tiancheng Wan, Bin Wang, Lei Xu and Xinlei Li

*Correspondence:
xiexin@ecjtu.edu.cn
East China Jiaotong University,
330013 Nanchang, China
Full list of author information is
available at the end of the article

Abstract

Most of the current image edge detection methods rely on manually features to extract the edge, there are often false and missed detections when the image has adverse interference. The surface of mechanical parts is smooth, when taking photos in the industrial field, it is easy to have specular reflection and shadow at the same time, which will affect the edge detection results. In order to achieve excellent edge detection performance, we propose a semantic segmentation model based on encoder-decoder structure. It adopts joint learning strategy, using two decoders to process image decomposition task and segmentation task respectively, and sharing their parameters to eliminate the influence of illumination, so as to improve segmentation performance. In the training phase, the asymmetric convolution and BN fusion are combined to improve the detection efficiency. In addition, we built a gear part dataset for experimentation. The result shows that in the task of edge detection of mechanical parts affected by illumination, our method has better performance than classical method.

Keywords: Joint Learning; Semantic Segmentation; Asymmetric Convolution; BN Fusion; Edge Detection

Introduction

Common mechanical parts such as gears and slender shafts are widely used in military industry, aerospace, automobile industry and manufacturing industry [1]. In actual production, the machining precision of mechanical parts is affected by the precision of machine tools, thermal deformation, datum end face positioning, cutting vibration and other factors [2], and the precision of parts directly affects the working performance and service life of the machine and equipment [3]. Therefore, the factory inspection of parts dimensional precision is an extremely important link in industrial production.

With the development of computer vision and nondestructive testing and other related technologies, there are many efficient non-contact measurement methods for mechanical parts, in which the segmentation algorithm of part image is crucial. Xie X et al. [4] use the improved Roberts operator to extract the contour of the target, and then Zernike moments are used for sub-pixel positioning. At the same time using the Otsu method to automatically select the segmentation threshold, which achieves good detection efficiency and detection accuracy. Zhu G et al. [5] use Canny operator and bilinear interpolation to extract sub-pixel edge, and the least square method is used to fit the circle contour, and the measurement of the concentricity of the precision parts has obtained good results. Ofir N et al. [6] regards edge detection as a group of discrete curves to search for faint edges with noise interference, and

effectively detects these faint edges. These traditional methods mentioned above mostly extract the edge by analyzing the shape, texture, color and other features of the target image [7], then calculate the size of the parts. Generally, to build a model for specific applications, we not only need be familiar with the process of edge detection, but also rely on the manually designed extractor for feature extraction. In addition, we also need to have professional knowledge and parameter adjustment process, which can't be widely applied.

In recent years, the development of deep learning theory and practice provides a good reference for image edge detection and segmentation. Among them, semantic segmentation is a challenging problem. Changes in imaging conditions may have a negative impact on the segmentation process. These changes include shadows, reflections, light source colors and intensity. Since image segmentation is a process of grouping pixels based on visual or semantic characteristics [8], sharp changes in pixel values may lead to inaccurate segmentation. The online detection of the mechanical parts precision in industrial scenes, due to the movement of the parts and the specular reflection of the metal surface, the image quality is poor. The existing models are not ideal for image processing with large differences in light and dark, which greatly affects the segmentation precision and accuracy of the parts image.

Intrinsic image decomposition is the process of decomposing an image into reflectance components and illumination components. The reflectivity component is an inherent property of the object and will not change with light factors. Illumination components change continuously with light source factors, including specular reflection and shadow that affect image semantic segmentation. Therefore, it is more effective to use the reflectance image for semantic segmentation because it does not contain the negative effects of illumination. On the contrary, the category information of semantic segmentation contains prior knowledge of object reflectivity, which can guide the intrinsic image decomposition.

We designed a convolutional neural network with an encoder-decoder structure. After extracting image features through one encoder, two decoders are used to process image segmentation and image decomposition tasks respectively. Regarding segmentation and decomposition as a mutually-promoting combined process, the model is trained with a joint learning strategy to eliminate the influence of light to improve the accuracy of mechanical parts segmentation.

The contributions of this paper are as follows:

- a) Design a semantic segmentation model for parts image, use the joint learning method to improve the semantic segmentation performance;
- b) Create a dataset of mechanical parts image with labels;
- c) Use asymmetric convolution and BN fusion to optimize the model performance.

Related Works

Multi-Task Learning

Multi-Task learning is a method that combines multiple tasks and learn simultaneously to enhance the ability of model representation and generalization. Joint learning can be realized through the neural network approach. Currently, the main work can be divided into two categories: Parameter Sharing [9] and Tagging strategy [10].

There are three existing parameter sharing schemes, as shown in Figure 1: Hard sharing [11], Soft sharing [12] and Hierarchical sharing [13]. Hard Sharing is the most widely used sharing mechanism today, embedding data representations of multiple tasks into the same semantic space, and then extracting feature representations for each task using a specific layer. Hard sharing is easy to implement and is good for tasks with strong correlation, but it often performs poorly when encountering weakly related tasks. For the soft sharing, each task uses a single network for learning, and the network of each task can access information in the network corresponding to other tasks, such as eigenvalues, gradients, and so on. Although soft sharing mechanism is very flexible and doesn't need any assumptions about task dependencies, additional parameters are required for assigning each task a network. Hierarchical sharing is to do the simple tasks in shallow layers of the network and complex tasks in deep layers. Hierarchical sharing is more flexible than hard sharing, and requires fewer parameters than soft sharing. However, designing an efficient hierarchical structure for multiple tasks completely relies on expert experience.

Image segmentation is based on the category information of the object, and the category information contains the prior knowledge of the object's reflectivity attributes, so segmentation and decomposition can be used as a strong correlation task. In this paper, hard sharing mechanism is used to extract features with one encoder, and two decoders handle the corresponding tasks separately. Furthermore, the soft sharing is also combined between decoders to improve the performance of each task by sharing parameters.

Image Semantic Segmentation

Image segmentation is a process of assigning a label to each pixel in the image such that pixels with the same label are connected with respect to some visual or semantic property [8], then divide to different region. Semantic segmentation technology based on deep learning has greatly improved the performance of image edge detection. Long J et al. [14] propose a full convolutional neural network (FCN), which

replaces the full connection layer with the convolution layer in CNN, retains the spatial information of the target in the output, and realizes semantic segmentation by pixel level classification. Benefit from rich spatial information and large perception domain [15], many classic models such as U-Net [16], Mask R-CNN [17], Deeplab [18], etc. are proposed based on FCN. Most of these models use public datasets such as ImageNet [19], to show the performance of model. According to the proposed model, researchers use some improvements and optimizations to satisfy the specific scenarios. In order to improve the segmentation performance, Stan T et al. [20] samples a large number of small images in a fixed number of X-ray dataset to train the neural network. Smith A et al. [21] design a U-Net based convolutional neural network and constructed an annotated chicory dataset, which successfully complete the root system segmentation task of plants. It also demonstrates the feasibility of using deep learning to create the own dataset. Vuola A et al. [22] compare the advantages and disadvantages between U-Net and Mask R-CNN, and develop an integrated model, which achieved better results in nuclear segmentation task.

In the task of mechanical parts image segmentation in industrial scene, the change of lighting conditions will cause the change of object appearance when the image is collected in the field, which has a negative impact on the semantic segmentation task. In this paper, we joint learning segmentation and decomposition tasks to reduce the impact of illumination during segmentation.

Intrinsic Image Decomposition

An image can be decomposed to generate countless combinations of reflectance and illumination, so image decomposition is a long-standing ill-posed problem [23]. Li et al. [24] add nonlocal texture constraints to traditional techniques to optimize intrinsic image decomposition, which is a significant improvement over previous algorithms. With the development of technology, the latest research on intrinsic image decomposition has turned to deep learning technology. Shi et al. [25] use neural network decoder to jointly optimize each component by learning the correlation between intrinsic attributes, and achieved robust and real decomposition result. Based on this study, we use image segmentation attributes as an assistant to improve the performance of other tasks in joint learning.

Method

In order to reduce the influence of illumination factors on parts image segmentation, we use joint learning method in the encoder-decoder model, a shared encoder is used to extract features, and two decoders are used to learn image segmentation and image decomposition respectively. Through intrinsic image decomposition, the reflectance image without illumination component is used to guide the semantic segmentation task; at the same time, the class attribute provided by semantic segmentation contains the prior knowledge of reflectance of the target object, which in turn guides the image decomposition task. In addition, asymmetric convolution and BN fusion are used to enhance the feature learning ability and accelerate the operation speed respectively.

Joint Learning Method

Image intrinsic decomposition is based on Retinex theory [26], which decomposes an image into the product of a reflectance image and illumination image. The results obtained by image decomposition are not unique, and most of the current work is devoted to solving the ill-posed problem of intrinsic image decomposition. Supposing I is the original image, R is the reflectance image, S is the illumination image, and (x, y) are the pixel coordinates in the image, then the classical intrinsic image decomposition can be formulated as:

$$I(x, y) = R(x, y) S(x, y) \quad (1)$$

Since the reflectance image is an object's own property, it is not affected by any light. According to the ShapeNet [25] model, given an image I , the process of obtaining reflectance component R and illumination component S by image intrinsic decomposition can be understood as follows:

$$(R, S) = F(I, \theta) \quad (2)$$

θ contains all the parameters learned by the image intrinsic decomposition decoder. MSE is used to optimize each component of θ . Let R^* to be the ground truth parameter in the dataset, and R is the parameter learned by decoder, and $r = R^* - R$ is the learning difference. In order to obtain the most realistic reflectance image, minimize the following formula:

$$L_1(R^*, R) = \frac{1}{n} \sum_{i,j,c} r_{i,j,c}^2 - \frac{1}{2n^2} \left(\sum_{i,j,c} r_{i,j,c}^2 \right)^2 \quad (3)$$

where i, j are pixel coordinates; n is the total number of pixels; C is the RGB channel index of color image.

Image semantic segmentation task assigns a label based on what the pixel represents. Through a series of convolution and pooling operations, we can obtain a low-resolution multichannel feature map, which contains the characteristics of the relationship between the object and its environment. The feature map can provide the contextual semantic information of the segmented target in the entire image.

For a dataset with n classes and labels, a prediction is made of the probability that the pixels belong to each class, and the sum of these probabilities is 1:

$$\sum_{i=1}^n P((x, y) | i) = 1 \quad (4)$$

P represents the probability that a pixel with coordinates (x, y) belongs to class i . After sampling from a deep feature map, P can be predicted, and the biggest P can be selected as the label of the pixel. Gather the pixels with same label to generate a Mask, which divides the same class of region.

For our image semantic segmentation task, to reduce the calculation, the dataset has only two classes: conveyor background and part foreground. We use loss function

L_2 to calculate the number of wrong pixels in foreground and background prediction. The smaller of L_2 the better outputs, so minimize the following formula:

$$L_2 = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

where y_i is the class of predicted pixels (0 as background, 1 as foreground); $\hat{y}_i$ is ground truth; n is the total number of pixels in the image.

For the whole model, in order to achieve joint learning, we combine the loss function of segmentation decoder and decomposition decoder to jointly train the parameters, as shown below:

$$L = \gamma_1 L_1 + \gamma_2 L_2 \quad (6)$$

γ is a manually set coefficient, which represents the weight of the corresponding loss function. By optimizing L to get the best output of image segmentation, and impact of γ will be shown in the evaluation part.

Asymmetric Convolution

To achieve better segmentation performance, most ideas for model improvement are mainly focused on: 1) how to connect the layers [27]; 2) combining different layers to improve the learning quality [28]. The asymmetric convolution [29] use an improved scheme independent of the network structure. It doesn't increase computation and can fulfill the real-time requirements of image detection in industrial field. The structure of asymmetric convolution is shown in Figure 2.

If several 2D kernels with compatible sizes operate at the same stride on the same input to produce the output with the same resolution and their outputs can be summed, these kernels can be added at the corresponding positions to obtain an equivalent kernel producing the same output:

$$I * K^1 + I * K^2 + I * K^3 = I * (K^1 \oplus K^2 \oplus K^3) \quad (7)$$

Let the input I , the convolution kernel K , and $\oplus$ is the addition of kernel parameters at corresponding positions. Taking the convolution kernel of 3×1 as an example, there are M convolution kernels, and the output of the j convolution kernel K^j is at J_{th} channel, the value of a point P in the output can be expressed as follows:

$$P_{:,j} = \sum_{c=1}^C \sum_{a=1}^3 \sum_{b=1}^1 K_{a,b,c}^j W_{a,b,c} \quad (8)$$

Where W is the corresponding sliding window. If the points P output by the three convolution kernels correspond to the same sliding window, then the additivity of formula 7 holds (dark color in Fig. 2).

Figure 2 Asymmetric convolution

BN Fusion

Batch normalization [30] can accelerate the convergence speed of model training, make the model training process more stable, and avoid gradient explosion or gradient disappearance. Usually, neural networks are batch normalized after convolution, which requires two calculations. BN fusion combines these two steps into one.

For a convolution layer, the output is determined by the weights ω and the bias b : $X_l = \omega * X_{l-1} + b$. The batch normalization is shown in formula 9:

$$X_{norm} = \gamma * \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (9)$$

γ and β are trainable hyperparameters, which are iterated by back propagation. As a restore parameter, it retains the distribution of the original data to a certain extent. μ is the mean value of input X , σ^2 is the variance, and ε is a very small constant to avoid errors of dividing by zero.

The homogeneity of convolution allows subsequent BN operations and linear scaling to be integrated into the convolution layer with additional bias. Expand X in formula 8 and take the following deformation:

$$X_{norm} = \frac{\gamma * \omega}{\sqrt{\sigma^2 + \varepsilon}} * X_{l-1} + \beta + \gamma * \frac{b - \mu}{\sqrt{\sigma^2 + \varepsilon}} \quad (10)$$

Construct a new convolution kernel that new weights $\tilde{\omega} = \frac{\gamma * \omega}{\sqrt{\sigma^2 + \varepsilon}}$, new bias $\tilde{b} = \beta + \gamma * \frac{b - \mu}{\sqrt{\sigma^2 + \varepsilon}}$. Then output of the new kernel is same as the result of original Conv+BN, but only one calculation. After the model is trained, using BN fusion to speed the inference time.

Model Structure

The model used in this paper, both encoder and decoder are 5 layers. The encoder uses a convolutional kernel of 3*3 and stride of 2 to extract the features of each layer, and then uses the batch normalization to reduce the correlation between the layers. After BN operation, the rectified linear unit (ReLU) [31] is used as the activation function. The decoder uses the feature size symmetrical to the encoder for up sampling. The model is shown in Figure 3.

There is a mirror link between the encoder and the decoder. We use the copy and crop method in U-Net [16] to make upsampling process more accurate. There is parameter sharing between the two decoders, the feature values after ReLU activation are shared with each other, there are two reasons for this: (1)The reflectance image obtained by the intrinsic decomposition doesn't contain illumination components, so as to guide the segmentation process to reduce the incorrect segmentation caused by the difference of highlight and shadow. (2)The class information obtained by semantic segmentation contains prior knowledge about the reflectance of an object, which can guide the generation of more accurate reflectance images.

In the training phase, we use three convolution kernels of 3 * 1, 1 * 3 and 3 * 3, all of which use a 3 * 3 sliding window to match the existing square convolution kernel. The branches of the three convolution kernels are all Conv + BN operation. After the training is completed, they are fused into a standard square convolution kernel of size 3 * 3. This process doesn't require any additional hyperparameters. The output composition is as follows:

$$O = \begin{cases} (I * K_1 - \mu_1) \frac{\gamma_1}{\sigma_1} + \beta_1 \\ (I * K_2 - \mu_2) \frac{\gamma_2}{\sigma_2} + \beta_2 \\ (I * K_3 - \mu_3) \frac{\gamma_3}{\sigma_3} + \beta_3 \end{cases} \quad (11)$$

I is the input image, and the output O is the sum of three branches after convolution and batch normalization. μ , γ , σ and β are the parameters in BN operation.

The enhanced square convolution kernel contains BN fusion:

$$O = I * \left(\frac{\tilde{\gamma}_1}{\tilde{\sigma}_1} K_1 \oplus \frac{\tilde{\gamma}_2}{\tilde{\sigma}_2} K_2 \oplus \frac{\tilde{\gamma}_3}{\tilde{\sigma}_3} K_3 \right) + \beta \quad (12)$$

$\tilde{\gamma}$, $\tilde{\sigma}$ and $\tilde{\beta}$ are the parameters after BN fusion. Branches can be converted into standard convolution kernel by adding kernel parameters at the corresponding position.

Results and discussion

The experiment configuration is as follows: CPU is AMD R5 2600, GPU is NVIDIA GTX 1660Ti, RAM is 16GB. Experiments are programmed in the TensorFlow framework [32], the code running environment is Python3.8, and the deep learning environment is CUDA10.1 and cudnn7.6. We use RMSProp [33] optimizer to train the model, the learning rate is 0.1, the batch size is 2, and the model converges after training for 100 epoches.

Dataset

In this paper, we need to make effective semantic segmentation for the images collected on the conveyor belt in industrial production, so we construct a new dataset containing lots of metal parts images. Considering the different illumination in different areas of the conveyor belt, different sizes of gears occupy different pixels. A total of 600 original images are collected by manual setting in the area without specular reflection. After collecting the original image, add 5% Gaussian noise through Photoshop to obtain noise images. There are 1200 images in the dataset, and we extract the target area and crop the image to 320*320 resolution. We randomly select 1000 images as the training set, and the remaining 200 images are used as the test set.

Intrinsic image decomposition training requires labels that can't be labeled by hand. Therefore, we use the part image with ideal shooting results as the benchmark, and rendering these images with specified intensity of highlight. The original low-light images are taken as the ground truth reflectance attributes. As for semantic segmentation label, we manually define the black area as the foreground and the white area as the background. Example of the dataset is shown in Figure 4.

Evaluation Criteria

The proposed model performs two tasks simultaneously, but the main target is image segmentation. Therefore, for intrinsic image decomposition decoder, we quote parameters from the existing model ShapeNet [27], and then fine-tune with the industrial parts dataset, and use MSE to measure the decomposition effect.

For the image semantic segmentation task, we use the pixel accuracy to evaluate the prediction effect:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

TP (True Positive) is the number of pixels correctly predicted as the target, TN (True Negative) is the number of pixels correctly predicted as the background. On

the contrary, FP (False Positive) and FN (False Negative) are the pixels with wrong prediction.

In order to comprehensively analyze the segmentation performance and compare with the mainstream segmentation models, we also use IoU to evaluate the segmentation effect:

$$IoU = \frac{target \cap prediction}{target \cup prediction} \quad (14)$$

Experimental Results and Analysis

Experiment 1: effectiveness of parameter sharing in joint learning

This experiment mainly verifies the effectiveness of parameter sharing. We use the control variable method to test the following cases. Case 1: without joint learning, only the semantic segmentation decoder works. Case 2: the intrinsic image decomposition decoder shares one-way parameters to the semantic segmentation decoder, as an aid to the segmentation task. Case 3: two decoders pass parameters to each other to achieve joint learning. The experimental results are shown in Table 1.

Table 1 Parameter sharing in joint learning

/	Segmentation		Decomposition	
	PA	IoU	Reflectance	MSE
case1	0.903	0.766	-	
case2	0.927	0.787	0.0113	
case3	0.941	0.813	0.0089	

The data shows that only a single decomposition task has a general result, and one-way assistance can improve the accuracy of semantic segmentation. In case3 with joint learning, not only the segmentation result is better, but also the decomposition is improved. This is because the two tasks promote each other and improve the performance.

Experiment 2: weights of loss function

This experiment verifies the influence of loss function weights. Let the sum of γ_1 and γ_2 is 1, we analyze the weights of two tasks on different proportions, and how it affect the segmentation performance of the model. The experimental results are shown in Table 2.

Table 2 Influence of loss function weights

γ_1	γ_2	PA	IoU
0.1	0.9	0.927	0.792
0.3	0.7	0.941	0.813
0.5	0.5	0.892	0.766
0.6	0.4	0.825	0.694
0.7	0.3	0.735	0.586
0.9	0.1	0.516	0.367

It can be seen from the table, the segmentation performance is improved when decomposition loss function γ_1 increased. However, with the weight γ_1 become bigger and bigger, the evaluation index of segmentation decays rapidly. Therefore, according to the experimental results, $\gamma_1 = 0.3$ and $\gamma_2 = 0.7$ are selected as relatively optimal combinations.

Experiment 3: influence of asymmetric convolution

This experiment compares the effects of normal convolution kernel and asymmetric convolution blocks on the model. This experiment was carried out on the parameter configuration of first two experiments, and three groups of test results are selected, as shown in Table 3.

Table 3 Performance of different convolution kernels

/	Test1		Test2		Test3	
	PA	IoU	PA	IoU	PA	IoU
convention	0.941	0.813	0.938	0.809	0.940	0.811
new kernel	0.945	0.818	0.943	0.813	0.945	0.816

It can be seen in the data comparisons, conventional kernel enhanced by the asymmetric convolution blocks can improve the performance of segmentation, but only a limited increase in the pixel accuracy of about 0.5%.

Experiment 4: influence of BN fusion

On the basis of previous experiment, the convolution layer and BN layer are fused in this experiment. In order to verify the ability of BN fusion to accelerate the processing, we divide the test set images into 10 batches, and every batch has 100 images. We calculate the time required to process each batch under different conditions, as shown in Figure 5.

According to the figure, each batch of images enhanced by BN fusion is generally faster than baseline, and the average time is increased by about 4.5%.

Experiment 5: comparison with classical semantic segmentation model

We verify the effectiveness of our model in the segmentation of industrial part dataset. We chose two existing classical models: FCN [7] and U-Net [9]. The experimental results are shown in Table 4.

Table 4 Results of different models

Model	PA	IoU
SegNet	0.908	0.759
U-Net	0.911	0.769
Ours	0.945	0.818

The conclusion in Table 4 shows that compared with the traditional single task segmentation model, joint learning can significantly improve the segmentation performance of the part dataset after eliminating adverse interference. The comparison results are shown in Figure 6.

Image of first group has simple object, and the segmentation results of our method and comparative method are both excellent. But for second group, the metal part image has specular reflection and shadow at the same time, our method can eliminate these two disadvantageous interferences. In the contrast experiment, SegNet eliminates shadow interference, but the external contour has many gaps; U-Net has continuous contour, but the shadow is also judged as target. In order to demonstrate the ability to resist the highlight influence, the third group of images have obvious light details on the upper right of the target, our method correctly identified it as the target, but the two comparison algorithms have some deficiencies in

some extent. The last group of images is a complex part, in the group, although our method is obviously better than comparison experiment, there are still many details to be optimized.

Conclusion

Aiming at machine parts images that contain the influence of illumination, this paper studies the use of deep learning-based semantic segmentation methods for edge detection of mechanical parts images. Although classic CNN methods can learn some effects of illumination, it can't completely eliminate these disadvantages. Therefore, this paper proposes a joint learning method to guide semantic segmentation through the reflectance feature map without the illumination influence to improve its prediction performance.

The method proposed in this paper is aimed at the image of mechanical parts, but each image in the dataset has only one target part. In fact, in the industrial field, the collected image usually contains multiple targets, and even some parts overlap together. In the future, the follow-up work will be carried out in multi-object instance segmentation to make the algorithm more suitable for the images collected on the industrial scene.

Abbreviations

CNN: Convolutional Neural Networks; MSE: Mean Square Error; BN: Batch normalization; ReLU: Rectified Linear Unit; TP: True positive; FN: False negative; FP: False positive; TN: True negative; IoU: Intersection over Union

Acknowledgements

The authors acknowledged the anonymous reviewers and editors for their efforts in valuable comments and suggestions.

Author's contributions

All authors took part in the discussion of the work described in this paper. All authors read and approved the final manuscript.

Author's information

Xin Xie is a professor in the School of Information Engineering, East China Jiaotong University. His research interests are computer networks, information security and machine vision. Tiancheng Wan, Bin Wang, Lei Xu and Xinlei Li are graduate student of School of Information Engineering at East China Jiaotong University; their areas of interest are image processing and machine learning.

Funding

This work is supported by the National Natural Science Foundation of China, under Grant No. 61762037, and the Science and Technology Key Research and Development Program of Jiangxi Province, under Grant No. 20202BBEL53004

Availability of data and materials

Please contact authors for data requests.

Competing interests

The authors declare that they have no competing interests.

References

1. Zaigang Chen, Y.S.: Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth. *Engineering Failure Analysis* **18**(8), 2149–2164 (2018)
2. Fumi Takeoka, A.K. Masaharu Komori: High-precision measurement of an involute artefact: by a rolling method and comparison between measuring instruments. *Measurement Science and Technology* **20**(4), 1–12 (2009)
3. Zhi, S.: Research on visual measurement method of medium and small module gears and tooth pitch measurement technology (2019)
4. Xie X, X.M. Ge S: An improved industrial sub-pixel edge detection algorithm based on coarse and precise location. *Journal of Ambient Intelligence and Humanized Computing* **10**, 1–10 (2019)
5. Zhu G, Y.X. Pan S: Design of concentric measurement system for precise parts based on machine vision. *Transducer and Microsystem Technologies* **34**(04), 77–79 (2020)
6. Ofir N, A.S. Galun M: On detection of faint edges in noisy images[j]. *IEEE transactions on pattern analysis and machine intelligence*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(4), 894–908 (2019)
7. Duan R, L.Y. Li X: Summary of image edge detection. *Optical Technique* **31**(3), 415–419 (2005)

8. Ghosh S, D.I. Das N: Understanding deep learning techniques for image segmentation. *ACM Computing Surveys* **52**(4), 1–35 (2019)
9. Zheng S, L.D. Hao Y: Joint entity and relation extraction based on a hybrid neural network. *Neurocomputing* **257**, 59–66 (2017)
10. Zheng S, B.H. Wang F: Joint extraction of entities and relations based on a novel tagging scheme. *Association for Computational Linguistics* **1**, 1227–1236 (2017)
11. Subramanian S, B.Y. Trischler A: Learning general purpose distributed sentence representations via large scale multi-task learning, 1–16 (2018)
12. Ruder S, A.I. Bingel J: Latent multi-task architecture learning. *AAAI* **30**(01), 4822–4829 (2019)
13. Hashimoto K, T.Y. Xiong C: A joint many-task model: Growing a neural network for multiple nlp tasks, 1923–1933 (2017)
14. Long J, D.T. Shelhamer E: Fully convolutional networks for semantic segmentation, 3431–3440 (2015)
15. Krizhevsky A, H.G. Sutskever I: Imagenet classification with deep convolutional neural networks, 1097–1105 (2012)
16. Ronneberger O, B.T. Fischer P: U-net: Convolutional networks for biomedical image segmentation, 234–241 (2015)
17. Kaiming H, P.D. Georgia G: Mask r-cnn, 2961–2969 (2017)
18. Chen L C, K.I. Papandreou G: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* **40**(4), 834–848 (2018)
19. Deng J, S.R. Dong W: Imagenet: A large-scale hierarchical image database, 248–255 (2009)
20. Stan Tiberiu, V.P.W. Thompson Z T: Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning. *Materials Characterization* **160**, 110–119 (2020)
21. Smith A G, S.R. Petersen J: Segmentation of roots in soil with u-net. *Plant Methods* **16**(1), 13–27 (2020)
22. Vuola A O, K.J. Akram S U: Mask-rcnn and u-net ensemble for nuclei segmentation, 208–212 (2019)
23. Baslamisli A S, D.P. Groenestege T T: Joint learning of intrinsic images and semantic segmentation, 289–305 (2018)
24. Li S, S.L. Ping T: Intrinsic image decomposition with non-local texture cues (2008)
25. Shi J, S.H.Y.S.X. Dong Y: Learning non-lambertian object intrinsics across shapenet categories, 5844–5853 (2017)
26. E, L.E.H.: Lightness and retinex theory. *Journal of the Optical Society of America* **61**(1), 1–11 (1971)
27. Kaiming He, S.R. Xiangyu Zhang: Deep residual learning for image recognition, 770–778 (2016)
28. Christian Szegedy, S.I. Vincent Vanhoucke: Rethinking the inception architecture for computer vision, 2818–2826 (2016)
29. Ding X, D.G. Guo Y: Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks, 1911–1920 (2019)
30. Ioffe S, S.C.: Batch normalization: accelerating deep network training by reducing internal covariate shift, 448–456 (2015)
31. Xavier G, Y.B. Antoine B: Deep sparse rectifier neural networks. *Journal of Machine Learning Research* **15**, 315–323 (2011)
32. Google.: Tensorflow, an open-source machine learning framework for everyone (2016)
33. Zou F, J.Z. Li S: A sufficient condition for convergences of adam and rmsprop, 11127–11135 (2019)

Figure Title and Legend

Figure 1: Parametersharing. Different parameter sharing methods. Figure 2: Asymmetric convolution. The process of training and inference using asymmetric convolution. Figure 3: Model structure for joint learning. The neural network model used in this paper. Figure 4: Example of dataset. A sample of the original image, rendered image and label. Figure 5: Processing time of each batch. The time to process each batch with BN fusion.

Figures

Figure 1

Parameter sharing

Figure 2

Asymmetric convolution

Figure 3

Model structure for joint learning

Figure 4

Example of dataset. The first image is taken in the ideal environment, which achieves the best imaging result; second image is the foreground and background manually marked, which can be used as the label

of image semantic segmentation; last is the image with interference factors collected in the simulated industrial environment as the input of our model.

Figure 5

Processing time of each batch. The time to process each batch with BN fusion.

Figure 6

Processing time of each batch