Examination Paper Image Segmentation with Adversarial Network

Siquan Hu1,2*, Muliang Zhang3, Zhiguo Shi2 and Min Zhang3

1 Shunde Graduate School, University of Science and Technology Beijing, Foshan, China
2 School of Computer and Communication Engineering, University of Science and Technology (Beijing), Beijing, China
3 School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China
Email: husiquan@ustb.edu.cn

Abstract. Examination paper analysis is important for students to improve their learning efficiency. Traditional paper examination papers are difficult to sort out, which makes collecting mistakes for future review both time-consuming and laborious, either by handwriting or existed software tools. To easy the process, we propose a layout analysis method combined with the conditional generative adversarial network (CGAN). The traditional semantic segmentation structure is improved and used as a generator in the network, while a discriminator is designed to make the segmentation results more accurate. The motivation is that the discriminator can judge the authenticity of the image, so it can help reduce the unreasonable phenomenon in the semantic segmentation results of the generator. The experimental results show that by this method the examination paper image can be wellly split into various components, which provides convenience for further sorting and analysis of the examination paper images.

1. Introduction

For a long time, students have been seeking methods to improve learning efficiency. A smart analysis of examination papers can enable students to summarize and sort out knowledge points, find weak points accurately and achieve rewarding with less effort. Traditional paper examination papers are difficult to sort out, which makes collecting mistakes for future review a burden. Some students copy the questions wrongly answered in previous examinations on a notebook by handwritings, which is both time-consuming and laborious. OCR software tools could help, but their functions are very limited. The software’s automation processes work poorly on tasks such as dividing various question areas, recognizing special symbols, processing pictures and tables, where human intervention are highly needed. Powerful segmentation tools using deep learning may alleviate such burden.

As an important field of computer vision, semantic segmentation has been widely used in our daily life [1]. Traditional research often relied on the contour and edge information to achieve segmentation tasks [2, 3]. Nowadays, their efficiency or accuracy has fallen behind. In recent years, deep learning has made semantic segmentation grow with leaps and bounds. Many deep neural networks have performed well in this research area [4, 5].

An examination paper usually contains question numbers, questions, pictures, handwritten answers and background parts. The purpose of document layout analysis is to classify and locate the content in
the document, which can accelerate paper analysis perfectly. Liu [6] designed an adjustment algorithm related to character size to locate mathematical formulas in English PDF documents. Yi and Gao [7, 8] proposed some strategies to improve CNN to locate the tables and formulas in the document image. These studies have achieved excellent results, but the data is often based on electronic documents, with a more standard structure and less interference. In the 2017 ICDAR layout analysis competition, many scholars provided a variety of ideas to deal with medieval manuscripts [9]. Afterwards, Sofia [10] proposed to use the improved U-Net structure dhSegment to participate in this research and achieved relatively satisfactory results.

In this paper, we apply the document layout analysis on the examination paper images. The traditional semantic segmentation model is improved to perform the task of distinguishing various components from the images. At the same time, conditional generative adversarial networks [11, 12] (CGAN) are used to reduce unreasonable phenomena in the segmentation task. Our improved segmentation model is trained to learn the features of the sections to locate their positions in the examination paper accurately. The sample images will be divided into four to six kinds of components according to the structure. By assigning different colour information to each component, the model will learn this segmentation intention and be used on paper segmentation automatically. The experimental results show that our semantic segmentation method can segment the examination paper images more accurately, and lay a good foundation for further examination papers analysis.

2. Materials and Methods

2.1. Data
Since there is no public dataset for layout analysis of the examination paper, we built one by collecting some mathematics examination papers from junior high schools in Beijing and taking photos for them by mobile phones in different environments. These photos are processed by adjusting the resolution, tilt correction, binarization, and labelling different parts with different colours. Considering that the formulas and traditional characters in the mathematics examination papers may need to be separated in the subsequent recognition process, we label question numbers, traditional characters, formulas, pictures and background in five categories. Eighty percent of the samples are used to train the model, and the remaining twenty percent are used to test the model’s effectiveness.

To verify our model can be well applied in other datasets, we select CSG18 and CSG863 from the public dataset DIVA-HisDB [13] for experiments. The dataset contains some pictures of medieval manuscripts, whose content has been divided into four categories: comments, decorations, main text body and background. There are 40 manuscript pictures in each data set. We take 30 pictures to train the model and 10 pictures to test the model’s segmentation effects.

2.2. Image Semantic Segmentation
Image semantic segmentation is an important branch of computer vision and deep learning. It classifies each pixel in the image and determines the category to which each point belongs. This makes it widely used in document layout analysis tasks. Currently, there are many excellent image semantic segmentation networks. They often have deep convolutional neural networks to extract image features, and restore image resolution through operations such as unsampling, unpooling, and deconvolution. U-Net [14] is a well-known image semantic segmentation network, which is named because it resembles the letter U. Compared with other semantic segmentation networks, U-Net has few parameters, which makes it perform better in speed. Our research will improve the structure of U-Net.

2.3. Network Structure
We combine CGAN’s adversarial idea with semantic segmentation. CGAN contains two networks: one is the generator network, the other is the discriminator network. The generator is responsible for generating the picture, and it generates the picture through the condition label. The discriminator is
responsible for judging whether a picture is real or generated by the generator. In training, the generator will try to make the generated picture as close as possible to deceive the discriminator. The discriminator will try to distinguish the two pictures as much as possible, which forms a dynamic confrontation behavior. The reaction to the loss function is expressed as Formula 1, where $G$ is the generator, $D$ is the discriminator, $x$ is the real picture, and $y$ is the condition label. $D(x|y)$ represents the probability that the discriminator judges that the real picture is true, and $D(G(z|y))$ is the probability that the discriminator judges that the generated picture is true.

$$V_c(D, G) = E_x[\ln D(x|y)] + E_z[\ln (1 - D(G(z|y)))]$$  \hspace{1cm} (1)

In our model, the generator in CGAN is replaced with an improved U-NET structure, which means that the generator will perform the semantic segmentation task to generate segmentation result images. At the same time, the discriminator combines the conditional image and the image to be determined as input, which is specifically a form of splicing the RGB three channels of the two images into six channels. Through convolution, batch normalization, activation function, and other operations, the final probability result is output. The model structure is shown in figure 1.

Figure 1. Model structure.

2.4. Implementation Details

The batch normalization method proposed by Sergey [15] can accelerate the training speed in deep learning and improve the generalization ability of the network. And the activation function can enhance the expression ability of the neural network. In recent years, the ReLU function [16] and its improved types have been used in multi-layer neural networks. To further improve the ability of the activation function, PReLU [17] with only a few parameters added was proposed. We apply these excellent methods to our network structure.

In semantic segmentation, cross-entropy is often used as a loss function. However, the proportion of various types of images in our examination paper images is unbalanced. This function is often less effective for the problem of category imbalance. In SegGAN [18], the author proposed to use mean square error (MSE, also known as L2) as a loss function and achieved good results. But we noticed that both L1 and L2 losses have their shortcomings. The derivative of L1 loss at 0 is not unique, which may affect convergence. The L2 loss gradient is too large at a distance away from the center value and is not stable enough. Therefore, we choose to use a smooth L1 loss defined in equation (2).
\[ L_{\text{smooth}} = \begin{cases} 0.5(y-x)^2 & |y-x| < 1 \\ |y-x| - 0.5 & \text{otherwise} \end{cases} \] (2)

The structure diagram of our model gives a hint that the loss of the CGAN discriminator will also affect the two networks. The traditional CGAN loss function is used in the discriminator to alleviate the unreasonable phenomenon in the examination paper segmentation images. And the generator loss is defined in Formula 3, where \( \alpha \) and \( \beta \) are the weights of loss functions.

\[ L_G = \alpha L_{\text{smooth}} - \beta (\ln D(G(y))) \] (3)

3. Results

3.1. Experiment of Examination Papers

To verify the proposed CGAN can improve the accuracy of the semantic segmentation model, we compared the CGAN semantic segmentation model with another semantic segmentation model without CGAN. Specifically, the with-CGAN model is the model provided in figure 1, while no-CGAN is the traditional segmentation model which is the remained model with the discriminator removed in figure 1. As a common evaluation indicator for semantic segmentation models, average crossover rate (mIoU) reflects the performance of the segmentation network. The mIoU results are shown in table 1 for both models.

|          | Number | Character | Formula | Picture | mIoU  |
|----------|--------|-----------|---------|---------|-------|
| no-CGAN  | 64.3   | 94.0      | 67.8    | 71.0    | 74.3  |
| with-CGAN| 78.5   | 94.5      | 70.3    | 77.8    | 80.3  |

Three samples and their results are shown in figure 2, where in each sample, four parts from left to right are the input image, the output of no-CGAN, the output of with-CGAN, and the target image.

![Figure 2](image)

3.2. Experiment of Historical Manuscripts

Although our research objects are examination paper images, other kinds of documents may benefit from our model’s layout analysis ability. Therefore, further experiments are taken, where CSG18 and CSG863 in DIVA-HisDB are tested and compared with the previous reported research results. The results used for comparison are mainly from the ICDAR2017 layout analysis competition. The experimental results are shown in table 2.

And our CGAN model’s partial segmentation results for two samples are shown in figure 3, where three parts from left to right are the input image, the output image, and the target image.
Table 2. IoU values of the document layout analysis.

|                | CSG18 | CSG863 | Overall |
|----------------|-------|--------|---------|
| KFUPM          | 64.69 | 59.88  | 62.29   |
| IAIS           | 74.96 | 75.46  | 75.21   |
| BYU            | 87.72 | 86.42  | 87.07   |
| MindGarage-2   | 88.37 | 86.70  | 87.54   |
| Demokritos     | 90.69 | 89.36  | 90.03   |
| MindGarage-1   | 93.57 | 89.63  | 91.60   |
| dhSegment      | 92.80 | 90.50  | 91.65   |
| Ours CGAN model| 92.51 | 91.27  | 91.89   |
| NLPR           | 93.65 | 92.71  | 93.18   |

Figure 3. Document layout analysis segmentation results.

4. Discussion

From our early investigations, aside from the most tedious manual transcription process, existed applications also have fatal shortcomings in many ways. For example, manually selecting the topics one by one makes the operation very tedious, the wrong recognition of the formula makes the entire task difficult to handle, and the omission of special situations such as pictures or tables makes the layout confusing. Our research is dedicated to making these processes easier by automatically segmenting the examination paper images.

Our experiments have proved that applying semantic segmentation on examination paper images is feasible. It classifies various components in the examination paper successfully. The comparison of mIoU results proves that the introduction of CGAN's confrontation ideas can improve the accuracy of segmentation. Through the analysis of the result pictures, it is not difficult to see that the final output is more likely to produce mixed colours if the model without CGAN to determine the authenticity of the generated results. The effect of introducing the discriminator is to minimize this phenomenon so that the result is closer to the target image.

In the experiment of layout analysis for the medieval manuscripts, our model has achieved relatively good results. Although it can be seen from the table that the performance of the model has not reached the optimal level, the performance is still better than most other models. It proves that our model has good versatility and can be competent for various documents' layout analysis tasks.

5. Conclusion

In this paper, we are committed to studying how to improve the existing deep learning methods to form a semantic segmentation model for examination paper images. In particular, we have improved the U-Net model from various aspects such as model structure, activation function and loss function to make it perform well in the semantic segmentation of examination paper images. To deal with the unreasonable phenomena produced by the segmentation results, we introduce conditional generative adversarial networks to enhance the model performance. We tested the performance of our model in the public dataset DIVA-HisDB and achieved good results. Considering various problems that may be encountered in the actual situation, we use a dedicated data set to train the network for image segmentation of examination papers.
Through experimental results, we have proved that our deep neural network can complete this task. This frees students from the limitations of existing tools and automates the complex examination paper analysis process, laying the foundation for subsequent processing. Further efforts may be worthy for a fully examination paper automation tool including the accurate recognition of text and formulas, the reasonable classification of text and other content.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (Grant No. 61977005), the Sichuan Science and Technology Program (No. 2018GZDZX0034) and the Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB (BK19CF003).

References
[1] Guo Y, Liu Y, Georgiou T and Lew M S 2018 A review of semantic segmentation using deep neural networks International Journal of Multimedia Information Retrieval 7 (2) 87-93.
[2] Weinland D, Ronfard R and Boyer E 2011 A survey of vision-based methods for action representation, segmentation and recognition Comput. Vis. Image Underst. 115 (2) 224-241.
[3] Ilea D E and Whelan P F 2011 Image segmentation based on the integration of colour-texture descriptors—A review Pattern Recognition 44 (10) 2479-2501.
[4] Chen L, Zhu Y, Papandreou G, Schroff F and Adam H 2018 Encoder-decoder with atrous separable convolution for semantic image segmentation European Conference on Computer Vision pp 833-851.
[5] Liu Y, Chen K, Liu C, Qin Z, Luo Z and Wang J 2019 Structured knowledge distillation for semantic segmentation Computer Vision and Pattern Recognition pp 2604-2613.
[6] Liu C, Zuo L, Li X and Tian X 2015 An improved algorithm for identifying mathematical formulas in the images of PDF documents IEEE International Conference on Progress in Informatics and Computing pp 252-256.
[7] Yi X, Gao L, Liao Y, Zhang X, Liu R and Jiang Z 2017 CNN based page object detection in document images International Conference on Document Analysis and Recognition pp 230-235.
[8] Gao L, Yi X, Liao Y, Jiang Z and Tang Z 2017 A deep learning-based formula detection method for PDF documents International Conference on Document Analysis and Recognition pp 553-558.
[9] Simistira F, Bouillon M, Seuret M, Würsch M, Alberti M, Ingold R and Liwicki M 2017 Competition on layout analysis for challenging medieval manuscripts International Conference on Document Analysis and Recognition (ICDAR2017) pp 1361-1370.
[10] Oliveira S A, Seguin B and Kaplan F 2018 dhSegment: A generic deep-learning approach for document segmentation International Conference on Frontiers in Handwriting Recognition pp 7-12.
[11] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets Neural Information Processing Systems pp 2672-2680.
[12] Isola P, Zhu J, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial Networks Computer Vision and Pattern Recognition pp 5967-5976.
[13] Simistira F, Seuret M, Eichenberger N, Garz A, Liwicki M and Ingold R 2016 DIVA-HisDB: A precisely annotated large dataset of challenging medieval manuscripts International Conference on Frontiers in Handwriting Recognition pp 471-476.
[14] Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation Medical Image Computing and Computer Assisted Intervention pp 234-241.
[15] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift International Conference on Machine Learning pp 448-456.
[16] Glorot X, Bordes A and Bengio Y 2011 Deep Sparse Rectifier Neural Networks *International Conference on Artificial Intelligence and Statistics* pp 315-323.

[17] He K, Zhang X, Ren S and Sun J 2015 Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification *IEEE International Conference on Computer Vision* pp 1026-1034.

[18] Zhang X, Zhu X, Zhang X, Zhang N, Li P and Wang L 2018 SegGAN: semantic segmentation with generative adversarial network *IEEE International Conference on Multimedia Big Data* pp 1-5.