Research on Deep Learning Based Methods for Text Detection

Jiaxi Guo[1,a,*], Chuansheng Wu[1,b]

[1]University of Science and Technology Liaoning, Anshan, Liaoning, China
[a]3414324108@qq.com, [b]gykwcs@163.com
**Corresponding author*

Abstract: *This paper introduces a deep learning model training approach and a text detection method that harnesses the power of artificial intelligence, particularly in the fields of computer vision and deep learning. This method is well-suited for applications in optical character recognition (OCR) and other related scenarios. We detail the training process of the deep learning model for text detection, which involves the utilization of both single character segmentation and text line segmentation sub-networks. The trained model can effectively identify text areas in an image, and the prediction of single character segmentation and text line segmentation can be achieved simultaneously. By combining both text segmentation methods, we enhance the overall accuracy of text area detection.*

Keywords: *Deep Learning, Text Detection, Model Training*

1. Preface

With the development of deep learning technology, text detection based on deep learning models has been widely used in industry and academia, such as travel instant translation, paper document electronic, sign recognition, picture text review, etc., and to realize the detection of text in images, first of all, we need to determine the text area in the image[1].

2. Deep learning model training methods for text detection

A deep learning model training method for text detection, the method comprising: obtaining a deep learning model to be trained, wherein the deep learning model comprises a single character prediction network and a text line prediction network, the single character prediction network comprising a single character segmentation sub-network and a first character quantity prediction sub-network, and the text line prediction network comprising a text line segmentation sub-network and a second character quantity prediction sub-network; selecting a first category of sample data and the label data of the currently selected first class sample data; input the currently selected first class sample data into the deep learning model to obtain the prediction results of the currently selected first class sample data, in which the prediction results include the prediction results of the single character segmentation, the prediction value of the first number of characters, the prediction results of the text line segmentation, and the prediction value of the second number of characters; based on the prediction results and label data of the currently selected first class sample data, the prediction network includes the single character segmentation subnetwork and the first number of characters prediction subnetwork, the text line segmentation subnetwork and the second number of characters prediction subnetwork. According to the prediction results of the first sample data and labeling data, the training parameters of the deep learning model are adjusted to obtain the trained deep learning model[2].

A text detection method, comprising: obtaining data to be detected; inputting the data to be detected into a pre-trained deep learning model, obtaining a single character segmentation prediction result and a text line segmentation prediction result of the data to be detected, wherein the deep learning model is trained based on any of the deep learning model training methods for text detection herein; determining a text region in the data to be detected according to the single character segmentation prediction result and the text line segmentation prediction result of the data to be detected. According to the single character segmentation prediction results and text line segmentation prediction results of the data to be detected, the text region in the data to be detected is determined[3].

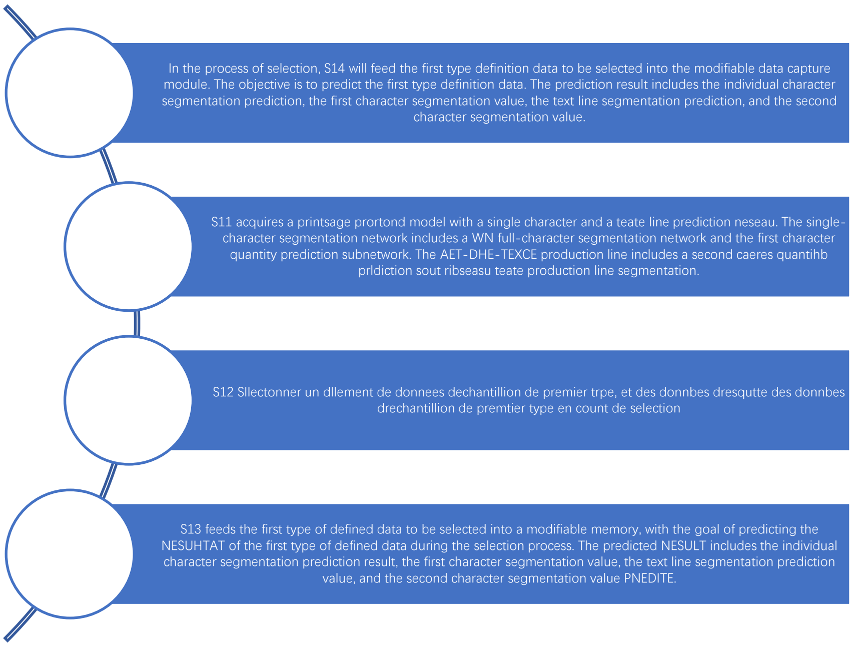
A non-transitory computer-readable storage medium storing computer instructions, wherein the computer instructions are used to cause the computer to perform a deep learning model training method for text detection or a text detection method of any one of the present applications.

A computer program product, comprising a computer program, the computer program, when executed by a processor, implements a deep learning model training method for text detection or a text detection method of any one of the present applications[4].

3. Deep learning model training device for text detection

A deep learning model training device for text detection, the device includes: a deep learning model acquisition module for obtaining a deep learning model to be trained, wherein the deep learning model includes a single character prediction network and a text line prediction network, the single character prediction network includes a single character segmentation sub-network and a first number of character prediction sub-network, the text line prediction network includes a text line segmentation sub-network and a second number of characters Prediction sub-network; first class sample data selection module, for selecting a first class sample data and the label data of the currently selected first class sample data; prediction result determination module, for inputting the currently selected first class sample data into a deep learning model to obtain a prediction result of the currently selected first class sample data, wherein the prediction result includes a single character segmentation prediction result, a first character number prediction result, a text line segmentation prediction result, a text line segmentation prediction result, a text line segmentation prediction result, a text line segmentation prediction result, and a text line prediction network. The predicted results include a single character segmentation prediction, a first character number prediction, a text line segmentation prediction, and a second character number prediction; the training parameter adjustment module is used to adjust the training parameters of the deep learning model according to the prediction results of the currently selected first class sample data and the labeling data to obtain a trained deep learning model[5].

A text detection device, comprising: a data to be detected acquisition module for obtaining data to be detected; a prediction result determination module for inputting the data to be detected into a pre-trained deep learning model to obtain single character segmentation prediction results and text line segmentation prediction results of the data to be detected, wherein the deep learning model is based on the deep learning model training device for text detection training obtained by either of the present applications The text region determination module is used to determine a text region in the data to be detected according to the single character segmentation prediction results and the text line segmentation prediction results of the data to be detected.



In the process of selection, S14 will feed the first type definition data to be selected into the modifiable data capture module. The objective is to predict the first type definition data. The prediction result includes the individual character segmentation prediction, the first character segmentation value, the text line segmentation prediction, and the second character segmentation value.

S11 acquires a printsage prortond model with a single character and a teate line prediction neseau. The single-character segmentation network includes a WN full-character segmentation network and the first character quantity prediction subnetwork. The AET-DHE-TEXCE production line includes a second caeres quantihb prldiction sout ribseasu teate production line segmentation.

S12 Sllectonner un dllement de donnees dechantillion de premier trpe, et des donnbes dresqutte des donnbes drechantillion de premtier type en count de selection

S13 feeds the first type of defined data to be selected into a modifiable memory, with the goal of predicting the NESUHTAT of the first type of defined data during the selection process. The predicted NESULT includes the individual character segmentation prediction result, the first character segmentation value, the text line segmentation prediction value, and the second character segmentation value PNEDITE.

Fig. 1 Schematic of the deep learning model training method for text detection

4. Specific implementation methods

In order to detect the text in an image, the text area in the image must be determined first. In view of this, the deep learning model training method for text detection in this embodiment, as shown in Figure 1, includes:

S11, acquire the deep learning model to be trained, wherein, the deep learning model includes single character prediction network and text line prediction network, single character prediction network includes single character segmentation sub network and first character quantity prediction sub network, and text line prediction network includes text line segmentation sub network and second character quantity prediction sub network[6].

The deep learning model training method for text detection in this embodiment can be realized by electronic devices, specifically, electronic devices can be smart phones, personal computers or servers[7].

The deep learning model to be trained includes a single-character prediction network and a text line prediction network, wherein the single-character prediction network includes a single-character segmentation subnet and a first character number prediction subnet, and the text line prediction network includes a text line segmentation subnet and a second character number prediction subnet. The single-character segmentation sub-network is used to predict the result of single-character segmentation, that is, to predict the region of each single character in the image; The text line segmentation sub-network is used to predict the result of text line segmentation, that is, to predict the area of each text line in the image; The first character number prediction sub-network and the second character number prediction sub-network are both used to predict the character number value, that is, to predict how many characters are in the image[8].

The specific network structures of the single-character segmentation sub-network, the first character number prediction sub-network, the text line segmentation sub-network and the second character number prediction sub-network can be customized according to the actual situation. In one example, the single-character segmentation sub-network can include multiple convolution layers and classifiers. The first character number prediction sub-network may include multiple convolution layers and fully connected layers; The text line segmentation sub-network can include multiple convolution layers, classifiers, etc. The second character number prediction sub-network may include a plurality of convolution layers and a fully connected laye[9]r.

S12, select a type I sample data and the label data of the currently selected type I sample data.

In one example, an unselected first-class sample data can be selected as the currently selected first-class sample data in a sample set including a plurality of first-class sample data. The first kind of sample data can be images. The first type of sample data has tag data, and the tag data of the first type of sample data includes at least one of the character number truth value, the single-character segmentation truth value result and the text line segmentation truth value result of the first type of sample data. The label data of the first kind of sample data can be obtained by manual labeling[10].

S13, input the currently selected first type of sample data into the deep learning model to get the prediction results of the currently selected first type of sample data, where the prediction results include single character segmentation prediction results, first character number prediction values, text line segmentation prediction results, and second character number prediction values.

The first type of sample data currently selected is input into the deep learning model, and the single character segmentation sub-network of the deep learning model outputs the corresponding single character segmentation prediction, the first character number prediction sub-network outputs the corresponding first character number prediction, the text line segmentation sub-network outputs the corresponding text line segmentation prediction, and the second character number prediction sub-network outputs the corresponding second character number prediction. In one example, each sub-network in the deep learning model can correspond to a separate feature extraction network, the first type of sample data is first input to the feature extraction network, and then input to the corresponding sub-network after extracting the features; in one example, each sub-network can share a feature extraction network; in one example, a part of the sub-network can share a feature extraction network, and a part of the sub-network corresponds to a separate feature extraction network, and a part of the sub-network corresponds to a separate feature extraction network, and a part of the sub-network corresponds to a separate feature extraction network. In an example, some sub-networks may share a feature extraction network, and some sub-networks may have a separate feature extraction network, all of which are within the scope of protection of the present application.

S14, according to the prediction results and tag data of the first type of sample data currently selected, adjust the training parameters of the deep learning model to obtain the trained deep learning model.

In one example, the loss of each network can be calculated separately based on the prediction results of the first type of sample data and the truth value in the labeled data, and the training parameters of the network can be adjusted according to the loss of the network, thus realizing the adjustment of the training parameters of the deep learning model.

For example, based on the single character segmentation prediction results and the single character segmentation truth value results of the first type of sample data currently selected, the first loss is calculated, and based on the first loss, the training parameters of the single character segmentation subnetwork are adjusted. For example, according to the predicted value of the first number of characters and the true value of the number of characters of the first sample data currently selected, the second loss is calculated, and according to the second loss, the training parameters of the first character number prediction subnetwork are adjusted. For example, according to the prediction result of text line segmentation and the true value of text line segmentation of the first sample data, calculate the third loss, and adjust the training parameters of the text line segmentation sub-network according to the third loss. For example, according to the predicted value of the second character number and the true value of the character number of the first sample data, the fourth loss is calculated, and the training parameters of the second character number prediction sub-network are adjusted according to the fourth loss.

For the method of adjusting the training parameters according to the loss, refer to the training parameter adjustment method in the existing technology. In one example, the training parameters of the network can be adjusted according to the SGD (Stochastic Gradient Descent) algorithm according to the loss.

After the completion of a training, continue to select the first type of sample data to train the deep learning model, until the preset training end conditions are met, and the trained deep learning model is obtained. The preset training end conditions can be customized according to the actual situation, such as the loss of the deep learning model convergence, or to reach the predicted number of training times. When the preset training end conditions are met, the training is stopped and the trained deep learning model is obtained.

5. Deep learning model training method implementation for text detection

In the embodiments of this paper, a deep learning model training method for text detection is given, and the trained deep learning model can be used for the detection of text regions; and the prediction of single character segmentation and text line segmentation can be realized at the same time, so that two text segmentation modes can be combined to carry out text detection, and the accuracy of the text region detection can be further improved.

In one possible implementation, the deep learning model further includes an encoder network, a first decoder network, and a second decoder network; the currently selected first class of sample data is inputted into the deep learning model to obtain a prediction result of the currently selected first class of sample data, comprising.

S21, use the encoder network to extract the features of the first type of sample data currently selected to obtain the global features.

In one example, the encoder network can use the lightweight Mobile-v3 network, combined with the Unet network, to extract the global features of the input image data and obtain the global features.

S22, use the first decoder network to extract the global features and obtain the first high-level features.

In one example, the first decoder network may include a multi-layer fully convolutional network for further feature extraction of the global features of the encoder network, and the resulting image features are referred to as the first high-level features. Wherein, the high-level features are image features with semantic information-rich target location roughness.

S23, use the second decoder network to extract the global features and obtain the second high-level features.

In one example, the second decoder network may include a multi-layer fully convolutional network for further high-level feature extraction of the global features of the encoder network.

S24, use the single character segmentation sub network to process the first high-level features to obtain the output single character segmentation prediction results, and use the first character number prediction sub network to process the first high-level features to obtain the first character number prediction value;

In one example, the first high-level features output from the first decoder network are passed through multiple convolutional layers in a single character segmentation subnetwork to obtain a feature map for single character pre-background classification, and then a single output map is obtained through a convolutional layer of a filter in the single character segmentation subnetwork to characterize the segmentation of the foreground and the background, and a single character segmentation prediction result with foreground of 1 and background of 0 is obtained. The first decoder network outputs the first high-level features through the first character number prediction sub-network of multiple convolutional layers for further feature extraction, and then through the first character number prediction sub-network of the full connectivity layer, the number of words prediction task as a classification task for prediction of the first character number prediction value, an example of a full connectivity layer of the output can be 1000 classes, corresponding to 0-999 characters, respectively. 999 characters.

S25, use the text line segmentation sub network to process the second high-level features, get the text line segmentation prediction result, use the second character number prediction sub network to process the second high-level features, get the second character number prediction value.

The second high-level features output from the second decoder network are passed through multiple convolutional layers in the text line segmentation sub-network to obtain the feature map for the pre-background classification of the text line, and then a single output map is obtained through the convolutional layer of the filter in the text line segmentation sub-network to characterize the segmentation of the foreground and the background, and to obtain the prediction result of the segmentation of the text line with foreground as 1 and background as 0. The second decoder network outputs the second high-level features through the second character number prediction sub-network of multiple convolutional layers for further feature extraction, and then through the second character number prediction sub-network of the fully-connected layer, the number of text prediction task as a classification task for prediction of the second character number prediction value, an example of the fully-connected layer of the output can be 1000 classes, corresponding to 0-999 characters, respectively. 999 characters.

The first high-level features extracted by the first decoder network are used for the prediction of single character prediction network, and the second high-level features extracted by the second decoder network are used for the prediction of text line prediction network. The training parameters of the first decoder network and the second decoder network can be adjusted separately, so that the decoupling of single character prediction network and text line prediction network can be realized to increase the recognition accuracy of single character prediction network and text line prediction network, thus ultimately improving the accuracy of text area detection and character number prediction. The accuracy of single character prediction network and text line prediction network can be increased, and the accuracy of text area detection and character number prediction can be improved.

6. Steps for tuning the training parameters of a deep learning model

In a possible embodiment, the labeling data of the first type of sample data includes at least one of a true value of a number of characters, a true value result of a single character segmentation, and a true value result of a text line segmentation; the step of adjusting the training parameters of the deep learning model based on the prediction results of the first type of sample data and the labeling data of the currently selected first type of sample data, comprising at least one of the following steps.

Step 1, calculate a first loss based on a single character segmentation prediction result of the currently selected first class of sample data and a single character segmentation truth result of the currently selected first class of sample data; according to the first loss, adjust the training parameters of at least one of the encoder network, the first decoder network, and the single character segmentation subnetwork.

Step 2, calculate a second loss based on the predicted value of the first number of characters of the currently selected first class of sample data and the true value of the number of characters of the currently selected first class of sample data; according to the second loss, adjust the training parameters of at least one of the encoder network, the first decoder network, and the first number of characters prediction subnetwork.

Step 3, calculate a third loss based on the prediction results of the text line segmentation of the currently selected first type of sample data and the true value results of the text line segmentation of the currently selected first type of sample data; according to the third loss, adjust the training parameters of at least one of the encoder network, the second decoder network, and the text line segmentation subnetwork.

Step 4, calculate a fourth loss based on the predicted value of the second number of characters of the currently selected first type of sample data and the true value of the number of characters of the currently selected first type of sample data; according to the fourth loss, adjust the training parameters of at least one of the encoder network, the second decoder network, and the second number of characters prediction subnetwork.

The first loss and the third loss can be a cross-entropy loss, for example, a binary cross-entropy loss. In one example, the predicted number of characters can be used as categories, for example, 1000 categories can be set, corresponding to the number of characters from 0 to 999, in which case the second and fourth losses can also be set as cross-entropy losses.

In the embodiment of this paper, the adjustment method of the training parameters of each network is given, which utilizes multiple losses to achieve the adjustment of the training parameters of each network, and can increase the accuracy of the prediction of each network.

In one possible embodiment, the method further comprises.

Step A, based on the first character number prediction value and the second character number prediction value of multiple first type sample data, determine the relative entropy of the first character number prediction value and the second character number prediction value, and obtain the first relative entropy.

Step B, according to the first relative entropy, adjust the training parameters of at least one of the first character number prediction subnet and the second character number prediction subnet.

7. Use of the first character count prediction subnetwork and the second character count prediction subnetwork

In this example, the first character number prediction sub network and the second character number prediction sub network are designed for DML (Deep Mutual Learning). KL divergence (relative entropy) is used to measure whether the predictions of these two sub networks match, and then training is conducted with the goal of restricting the matching degree of the two sub networks. This is because the input feature training of the first character number prediction sub network involves the single character position monitoring information, so it can more accurately predict the number of single characters. Let the two character quantity prediction subnetworks learn from each other, so that the prediction results of the second character quantity prediction subnetworks and the first character quantity prediction subnetworks are as consistent as possible, so that the second character quantity prediction subnetworks learn the knowledge of the first character quantity prediction subnetworks. Because the first character number prediction subnet and the second character number prediction subnet start training from different initial conditions and have different input characteristics, although they have the same label, their estimates of the probability of the next most likely class are different, and they learn from each other deeply, providing additional knowledge for training, This can further improve the accuracy of the prediction of the deep learning module, that is, the accuracy of the text detection.

In one possible implementation, obtaining the trained deep learning model, comprising.

Continue to select the first type of sample data for supervised training of the deep learning model, and use the second type of sample data for unsupervised training of the deep learning model, until it meets the preset training end conditions, to obtain the trained deep learning model.

Supervised training is the process of using the first type of sample data to train the deep learning model in the above embodiment. In one example, the sample data of each batch is composed of three parts. For example, the dimension of the sample data of a batch can be (3 * B, 3512512), which shows that 3 * B RGB (an image format) images with width times height of 512x512 are displayed, The first B images can be marked with single character annotation data (including the true value of the number of characters and the true value of the single character segmentation), the middle B images are marked with text line annotation data (including the true value of the number of characters and the true value of the text line segmentation), and the last B images are marked with unqualified text line annotation data. The

3 * B sheet here is the super parameter of model training, which is usually determined according to computing resources. When the sample data of a batch flows through the encoder network, the corresponding global characteristics are obtained. Next, the global characteristics pass through DecoderA (decoder A is the first decoder network) and DecoderB (decoder B is the second decoder network) at the same time, and the corresponding characteristics FA (first high-level characteristics) and FB (second high-level characteristics) are obtained. Feature FA then performs single character segmentation and total character prediction through single character prediction network to obtain single character segmentation prediction results and first character number prediction values; Feature FB performs text line segmentation and character total prediction through the text line prediction network, so as to obtain the text line segmentation prediction result and the second character number prediction value. Where, cross entropy represents cross entropy loss, and Binary cross entropy represents binary cross entropy loss. KL loss refers to KL divergence loss, and label refers to label.

When the deep learning model meets the first training conditions in the case of supervised training, join the unsupervised training and supervised training at the same time, an example of supervised training process can be shown by constraining the prediction of unlabeled samples before and after data augmentation is the same, to alleviate the problem of overfitting of the model, the relevant text detection technology, because it does not involve the prediction of the number of characters, the common data augmentation In related text detection techniques, because it does not involve the prediction of the number of characters, the commonly used data augmentation methods include cropping, etc. However, in the embodiment of this paper, the number of characters needs to be predicted, so in the embodiment of this paper, the data augmentation methods such as blurring, rotating, flipping, and stylization are used without changing the number of characters.

In the unsupervised training phase, the sample data of each batch is composed of two parts. Suppose the dimension of the sample data of a batch is (2 * N, 3512512), representing 2 * N RGB images with 512x512 width by height. The first N images are arbitrary sample images, and the next N images are the corresponding augmented data of the first N images. The augmented methods include at least one of blur, rotation, flip, and stylization. When the sample data of each batch passes through the encoder network, the global characteristics corresponding to the unlabeled data (equivalent to the second sample data) are input into decoder A, and then through the first character number prediction sub network, the character number prediction value of the non augmented sample data (equivalent to the third character number prediction value) is obtained. The global features corresponding to the unlabeled augmented data (equivalent to the third sample data) are input to decoder B, and then the character number prediction value of the sample data (equivalent to the fourth character number prediction value) is augmented through the second character number prediction subnet. Based on the predicted number of the third character and the predicted number of the fourth character, KL divergence is used to learn the consistency of the first character number prediction subnet and the second character number prediction subnet. In the process of unsupervised training, the single character segmentation sub network and text line segmentation sub network are not trained. Where KL loss represents KL divergence loss.

The first training condition can be set according to the actual situation, for example, the number of training times reaches the preset first training times, or the convergence degree of the deep learning model reaches the first convergence degree, etc. The preset training end condition can be set according to the actual situation. The preset training end conditions can be set according to the actual situation, for example, the number of training times reaches the preset second training times, or the convergence of the deep learning model reaches the second degree of convergence. Among them, the preset first training number is smaller than the preset second training number, and the convergence range of the first convergence degree is larger than the convergence range of the second convergence degree.

8. Examples of unsupervised training processes

The following is an example of an unsupervised training process, in one possible embodiment, unsupervised training of a deep learning model utilizing a second class of sample data, comprising.

Step A, obtain multiple type II sample data.

Step B, respectively expand the data of each second type of sample data to obtain the third type of sample data corresponding to each second type of sample data.

Step C, input the second type of sample data into the depth learning model after training, and obtain the third character number prediction value of the second type of sample data output by the first character number prediction sub network.

Step D, input each third type of sample data into the depth learning model after training, and obtain the fourth character number prediction value of each third type of sample data output by the second character number prediction sub network.

Step E, based on the predicted value of the third character number of each second type of sample data and the predicted value of the fourth character number of each third type of sample data, determine the relative entropy of the predicted value of the third character number and the predicted value of the fourth character number, and obtain the second relative entropy.

Step F, according to the second relative entropy, adjust the training parameters of at least one of the first character number prediction subnet and the second character number prediction subnet.

In the embodiments of this paper, the deep learning model is trained by supervised training and unsupervised training in two ways, and different learning tasks are combined for different data, and the training logic is simple. In the unsupervised training process can make full use of the massive unlabeled sample data for consistency learning, can reduce the model overfitting, and the use of unlabeled sample data for model training, can ensure the final text detection accuracy under the premise of reducing the workload of the sample data labeling, can be applied to the scene of the labeling of less data.

S51, obtain the data to be tested. The data to be detected can be any image data containing characters.

S52, input the data to be detected into the pre trained deep learning model, and obtain the single character segmentation prediction results and text line segmentation prediction results of the data to be detected.

Among them, the training process of the deep learning model can be referred to the deep learning model training method for text detection in the above embodiment, and the structure of the deep learning model can be referred to the structure of the deep learning model in the above embodiment, which will not be repeated here.

In a possible embodiment, the deep learning model is a deep learning model in which the first character number prediction subnetwork and the second character number prediction subnetwork are removed. In the text detection stage, the first character number prediction subnetwork and the second character number prediction subnetwork in the deep learning model can be removed on the basis of the deep learning model structure in the above embodiment, thereby reducing the data volume of the deep learning model and saving the operation resources of the first character number prediction subnetwork and the second character number prediction subnetwork.

S53, according to the single character segmentation prediction result and text line segmentation prediction result of the data to be detected, determine the text area in the data to be detected.

On the basis of single character segmentation prediction results and text line segmentation prediction results, the phase or operation of the text region is performed, and then the peripheral contour of the connected region is taken as the contour of the final detected text region.

In the embodiment of this paper, text detection is realized, using deep learning models to simultaneously realize the prediction of single character segmentation and text line segmentation, combining two text segmentation methods for text regions, which can improve the accuracy of text region detection.

In a possible embodiment, a text region in the data to be detected is determined based on a single character segmentation prediction result and a text line segmentation prediction result of the data to be detected, comprising.

S61, according to the single character segmentation prediction results of the data to be detected, the regions with predicted characters in the data to be detected are marked as the first value, and the regions without characters are marked as the second data to obtain the first binary map.

S62, according to the text line segmentation prediction results of the data to be detected, the regions with predicted characters in the data to be detected are marked as the first value, and the regions without characters are marked as the second data to obtain the second second value map.

S63, union the area of the first value in the first binary image with the area of the first value in the second binary image to obtain the text area of the data to be detected.

The first binary map in the first value of the region and the second binary map in the first value of the region to take the concatenation, and take the concatenation of the peripheral contour of the connected region that is ultimately detected as the contour of the text area.

In the embodiment of this paper, text detection is realized, and the merging of single character segmentation and text line segmentation can be realized accurately and efficiently by means of binary value map, which increases the detection efficiency of the text region and improves the accuracy of the text detection region.

Embodiments herein are also used for a deep learning model training device for text detection, the device includes: a deep learning model acquisition module 701, for acquiring a deep learning model to be trained, wherein the deep learning model includes a single character prediction network and a text line prediction network, the single character prediction network includes a single character segmentation subnetwork and a first character quantity prediction subnetwork, the text line prediction network includes The prediction result includes a single character segmentation prediction result, a first character number prediction value, a text line segmentation prediction result, and a second character number prediction value; a training parameter adjustment module 704 is used for obtaining a trained deep learning model based on the prediction result and the labeled data of the currently selected first class sample data.

The first high-level feature extraction submodule is used to use the first decoder network to extract features from the global features to obtain the first high-level features; the second high-level feature extraction submodule is used to use the second decoder network to extract features from the global features to obtain the second high-level features; the first prediction submodule is used to use the single character segmentation sub-network to process the first high-level features to obtain the output single character segmentation prediction results, and use the first character prediction sub-network to process the first high-level features, and use the first character prediction sub-network to obtain the first character prediction results. The first prediction sub-module is used to process the first high-level features by using the text line segmentation sub-network, to obtain the output text line segmentation prediction result, and to process the second high-level features by using the second character quantity prediction sub-network, to obtain the second character quantity prediction value.

In a possible embodiment, the device further comprises: a mutual learning module for determining the relative entropy of the predicted value of the first character number and the predicted value of the second character number based on the predicted value of the first character number and the predicted value of the second character number from the first sample data of the first class, obtaining a first relative entropy; adjusting the training parameters of at least one network in the prediction subnetwork of the first character number and the prediction subnetwork of the second character number based on the first relative entropy; adjusting the training parameters of at least one network in the prediction subnetwork of the first character number and the prediction subnetwork of the second character number. According to the first relative entropy, adjust the training parameters of at least one of the first character number prediction subnetwork and the second character number prediction subnetwork.

In a possible implementation, the deep learning model training module, specifically for: continuing to select a first class of sample data for supervised training of the deep learning model, and utilizing a second class of sample data for unsupervised training of the deep learning model until a preset training end condition is met, obtaining the trained deep learning model.

In a possible embodiment, the deep learning model training module is specifically used to: obtain a plurality of second class sample data; respectively, perform data augmentation on the second class sample data to obtain the third class sample data corresponding to the second class sample data; respectively, input the second class sample data into the trained deep learning model to obtain the third character number prediction value of the first character number prediction subnetwork output of the second class sample data; respectively, input the third class sample data into the trained deep learning model to obtain the fourth character number prediction value of the second character number prediction subnetwork output of the third class sample data; respectively, input the second character number prediction subnetwork output of the fourth character number prediction value of the third class sample data. The predicted value of the third character number is input into the trained deep learning model, and the predicted value of the fourth character number is obtained from the second character number prediction subnetwork; based on the predicted value of the third character number of the second sample data and the predicted value of the fourth character number of the third sample data, the relative entropy of the predicted value of the third character number and the predicted value of the fourth character number are determined, and the relative entropy of the predicted value of the third character number and the predicted value of the fourth character number are determined. Based on the predicted value of the third character number of each second class sample data and the predicted value of the fourth character number of each third class sample data, determine the relative entropy of the predicted value of the third character number and the predicted value of the fourth character number to obtain the second relative entropy; according to the second relative entropy, adjust the training parameters of at least one of the first character number prediction subnetwork and the second character number prediction subnetwork.

Embodiments of this paper also include a text detection device, comprising: a to-be-detected data acquisition module for acquiring to-be-detected data; a prediction result determination module for inputting the to-be-detected data into a pre-trained deep learning model, obtaining a single character segmentation prediction result of the to-be-detected data and a text line segmentation prediction result, wherein the deep learning model is based on the deep learning model training of any one of the deep learning models used for text detection in this application; a text region determination module for determining a text region in the to-be-detected data according to the single character segmentation prediction result and the text line segmentation prediction result. wherein the deep learning model is trained based on any of the deep learning model training devices for text detection in the present application; a text region determination module for determining a text region in the data to be detected based on the single character segmentation prediction results and the text line segmentation prediction results of the data to be detected.

According to the prediction results of the text line segmentation of the data to be detected, the region of the predicted characters in the data to be detected is labeled as the first numerical value, and the region without characters is labeled as the second data to obtain the second binary map; the first numerical value of the region in the first binary map is merged with the first numerical value of the region in the second binary map to obtain the text region of the data to be detected.

In one possible implementation, the deep learning model is a deep learning model that removes the first character number prediction subnetwork and the second character number prediction subnetwork.

The acquisition, storage and application of users' personal information in the technical solutions herein are in compliance with relevant laws and regulations and do not violate public order and morals.

According to embodiments herein, there are also electronic devices, a readable storage medium, and a computer program product herein.

An electronic device, comprising: at least one processor; and a memory communicatively coupled to the at least one processor; wherein the memory stores instructions executable by the at least one processor, the instructions being executed by the at least one processor to enable the at least one processor to execute a deep learning model training method for text detection or a text detection method of any one of the present applications.

A non-transitory computer-readable storage medium storing computer instructions, wherein the computer instructions are used to cause the computer to perform a deep learning model training method for text detection or a text detection method of any one of the present applications.

A computer program product, comprising a computer program, the computer program, when executed by a processor, implementing a deep learning model training method for text detection or a text detection method of any one of the present applications.

9. Conclusion

A deep learning model training method for text detection is given, and the trained deep learning model can be used for text region detection; and the prediction of single character segmentation and text line segmentation can be realized at the same time, so as to be able to combine two kinds of text segmentation methods for text detection, which can further improve the accuracy of text region detection. It should be understood that the contents described in this part are not intended to identify the key or important features of the embodiments herein, and are not intended to limit the scope of this paper. Other features herein will be readily understood by the following specification.

References

[1] Mnih, V. (Volodymyr), Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. A. (2013). Playing Atari with Deep Reinforcement Learning. CoRR, 89-90.

[2] Xie, E., Zang, Y., Shao, S., Yu, G., Yao, C., & Li, G. (2019). Scene Text Detection with Supervised Pyramid Context Network. Proceedings of the AAAI Conference on Artificial Intelligence, 56-57.

[3] Liu, Y., Jin, L., Zhang, S., Luo, C., & Zhang, S. (2019). Curved scene text detection via transverse and longitudinal sequence connection. Pattern Recognition, 103-108.

[4] Liao, M., Shi, B., & Bai, X. (2018). TextBoxes++: A Single-Shot Oriented Scene Text Detector. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 23-25.

[5] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 89-90.

[6] Toledo, R. Y., Caballero Mota, Y., & Martínez, L. (2018). A Recommender System for Programming Online Judges Using Fuzzy Information Modeling. Informatics, 77-78.

[7] Yera Toledo, R., Caballero Mota, Y., & Martínez, L. (2020). An e-Learning Collaborative Filtering Approach to Suggest Problems to Solve in Programming Online Judges. International Journal of Distance Education Technologies (IJDET), 66-67.

[8] Yera, R., & Martínez, L. (2017). A recommendation approach for programming online judges supported by data preprocessing techniques. Applied Intelligence, 54-56.

[9] De La Torre, J. (2009). DINA model and parameter estimation: A didactic. Journal of educational and behavioral statistics, 34(1), 115-130.

[10] Fan, X. (2012). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. Educational and Psychological Measurement, 67-68.