International Journal of Science and Engineering Applications
Volume 13-Issue 12, 28 – 31, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1312.1007

A Convolutional Long Short-Term Memory-Based method for labeled particle trajectory detection

Yichao Tan
School Electronic Information
and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: The traditional method of obtaining the trajectories of labeled particles requires a lot of manual operations, such as manually selecting a particle in each image and obtaining its center, and then obtaining its trajectory by connecting the lines, which not only can only obtain the trajectory of one labeled particle, but also needs a lot of time to get the center of the labeled particles. Aiming at the above problems, a labeled particle trajectory detection method based on Convolutional Long Short-Term Memory is proposed for the first time. First, a large number of consecutive frames are captured by an industrial camera and preprocessed using an algorithm, and a dataset is constructed using the preprocessed images. Then, a Convolutional Long Short-Term Memory network is constructed and trained on the dataset using this network. Finally, the trained model is tested using a test set and evaluated by metrics. The test results show that the PSNR of the model on the test set is 40.44, the SSIM is 0.95, and the LPIPS is 0.14, and all these figures indicate that Convolutional Long Short-Term Memory has achieved success in acquiring labeled particle trajectories.

Keywords: labeled particles; Motion trajectory; Image detection; Convolutional Long Term Short Term Memory Network

1. INTRODUCTION

In today's industrial production, granular substances are commonly used as basic materials in industrial production, and they play a key role in a number of industries, including the construction industry, metal smelting, chemical industry, material science, and food processing. The behavior of these granular substances during the manufacturing process plays a significant role in the quality of the final product. In order to gain a deeper understanding of the kinematic properties of powdered substances, many researchers have introduced labeling techniques to track their paths through manufacturing facilities and to monitor the paths these particles take. The technical tools used for tracking include a variety of techniques such as positron emission tomography (PET)[1], magnetic resonance imaging (MRI)[2], and visual inspection[3].

Compared with other methods, image detection method, as an advanced, low-cost and efficient technical means, has been widely used in many fields, and its core advantage is that it can provide intuitive and fast image analysis, and through the breakthroughs of deep learning technology, especially convolutional neural network, it makes the detection accuracy and efficiency increase dramatically. Currently, the image detection method is able to handle large-scale datasets and save resources by migrating models between different tasks with the help of transfer learning technology. In addition, the application of image detection method in augmented reality and virtual reality is also increasing, and the semantic relevance of image retrieval is also increasing through the development of natural language processing technology, which enables users to retrieve images through natural language query, improves the intuition of experiments and the clarity of images, and at the same time reduces the cost and improves the efficiency, which makes the image detection method in modern scientific research and industrial applications occupy an important position in modern scientific research and industrial applications.

Image detection methods can be categorized into traditional image detection methods and deep learning image detection methods. The main difference between traditional image detection methods and deep learning image detection methods is the way of feature extraction and the learning ability of the model. Traditional image detection methods like e.g. SIFT, SURF, HOG, etc., rely on hand-designed image features such as color, texture, shape, etc., which require expert knowledge and a lot of engineering time to determine, and are more adaptive to a specific scene, with limited generalization capability. In contrast, deep learning image detection methods automatically learn feature representations from a large amount of data through multi-layer neural networks, and are able to deal with more complex scenes and variations with stronger generalization ability and robustness. Deep learning methods are able to automatically extract rich information from images and adapt to various changes in the background and target, but require a large amount of labeled data and computational resources for training. Overall, deep learning image detection methods have obvious advantages in feature learning capability and model optimization, but the cost and resource requirements are high compared to traditional image detection methods.

However, Convolutional Long Short-Term Memory has generally been used in the field of video prediction in between, and this article will explore the feasibility of the model for the application of detecting marker particle trajectories.

2. Dataset production

This section focuses on the equipment used for the experiment and how the dataset was created.

2.1 Experimental equipment

In this study, an experimental device was constructed for detecting the trajectory of particles in a rotating cylinder, as demonstrated in Fig. 1. The setup includes a rotary cylinder, a motor, an industrial camera, a light source system, and several support structures. The rotary cylinder is a semi-open iron cylindrical vessel with a diameter and depth of 250 mm and 35 mm, respectively. The cylinder was driven by a motor whose speed was regulated by a frequency conversion device and was set between 10 and 20 revolutions per minute in the experiments. The rotary cylinder was filled with 30% to 50% fill rate, and particles were put in to satisfy the fill rate, which consisted of both labeled and non-labeled particles of 6 mm in diameter, and the number of non-labeled particles was ten times that of the labeled particles. A transparent plastic plate was fixed to one end of the rotating cylinder to facilitate observation of the movement of the particles inside through the plate. An industrial camera was placed directly in front of the rotating cylinder to capture the image of the particles inside the cylinder at a speed of 30 frames per second with a resolution of 128×128 pixels. One frame is shown in Fig. 2.The distance between the camera and the rotating cylinder was 400 mm. In order to minimize the influence of ambient light and improve the image clarity, a stable light source was used for the experiments.

Figure.1 Experimental equipment

Figure.2 One frame taken by industrial camera

The computing platform used in this paper is configured as follows: the CPU is a 13th Gen Intel® CoreTM i7-13700F, the GPU is an NVIDIA GeForce RTX 4070Ti®; the deep learning framework chosen is pytorch-cuda 11.7, the programming language is python 3.11, and the IDE uses the pycharm 2023.2.1 (Community Edition).

2.2 Image Preprocessing

In order to improve the accuracy of the Convolutional Long Short-Term Memory model in detecting the motion trajectory of marker particles, it is necessary to carry out preprocessing operations on the input image, and the specific steps are as follows: firstly, the original image is processed by using binarization and filtering to obtain the mask image of the marker particles, and then do the operation of comparison with the mask image and the original image to remove the background and retain the foreground in the original image. Then, the image is grayed out so that the number of channels in the image is reduced to reduce the amount of computation. Finally, a sliding window of length 8 and step size 1 is used to traverse all the images, and the image selected each time through the sliding window is the input to the model. During model training, not only input images but also labels are required. The labels are synthesized using pixel maximum synthesis for the 8 images selected each time by the sliding window. One labeled image is demonstrated in Fig. 3. After the above operation, we get a set of sample data consisting of 8 images with their corresponding labels. Repeating the above operation, 12,593 samples can be obtained, of which 70% is used as a training set for model training, and the remaining part is used as a test set for testing and evaluating the model.

Figure.3 labeled image

3. Introduction to indicators

In this section, we will introduce in detail the three metrics, PSNR, SSIM, and LPIPS, which are capable of evaluating the quality of model-generated images in a holistic manner.

3.1 PSNR

PSNR(Peak Signal-to-Noise Ratio) [4] is a full-reference measure of image quality, which evaluates the reconstruction quality of an image by comparing the mean-square error (MSE) between the original image and the distorted image. PSNR is based on the global evaluation of pixel values without considering the structural information of the image, and the higher the value, the higher the quality of the image, and usually a value of PSNR above 30 dB indicates better image quality, and above 40 dB, the difference is difficult to distinguish with the naked eye. A PSNR value above 30 dB indicates better image quality, and above 40 dB it is difficult to distinguish the difference with the naked eye.

3.2 SSIM

SSIM (Structural Similarity Index) [5] is a kind of image quality evaluation index that conforms to the human visual system, which measures the image similarity from brightness, contrast, and structure. SSIM takes the value in the range of [-1,1], and the larger the value, the smaller the image distortion is, and its value in 1 means that the two images are exactly the same, and the closer it is to 1 means the higher the similarity is.

3.3 LPIPS

LPIPS (Learned Perceptual Image Patch Similarity) [6] is a deep learning-based image similarity metric that evaluates the quality of an image by learning the mapping of a generated image to a real image. LPIPS can more accurately reflect human visual perception, is applicable to a variety of image types, and can be widely used in computer vision tasks such

International Journal of Science and Engineering Applications
Volume 13-Issue 12, 28 – 31, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1312.1007

as image recognition, image generation and image restoration. LPIPS can more accurately reflect human visual perception and is suitable for various types of images, and can be widely used in computer vision tasks such as image recognition, image generation, and image restoration. LPIPS measures the similarity of two images by comparing their feature differences, and the smaller its value is, the more similar the two images are.

4. Model training and testing

This section mainly includes model training and model testing. The model training section will briefly introduce the model structure and model parameters, as well as some model details. The model testing section will test the model with test and go to test the model, and the results obtained will be evaluated by PSNR, SSIM and LPIPS.

4.1 Model training

First we need to construct the framework of the model, the overall framework of Convolutional Long Short-Term Memory is shown in Fig. 4

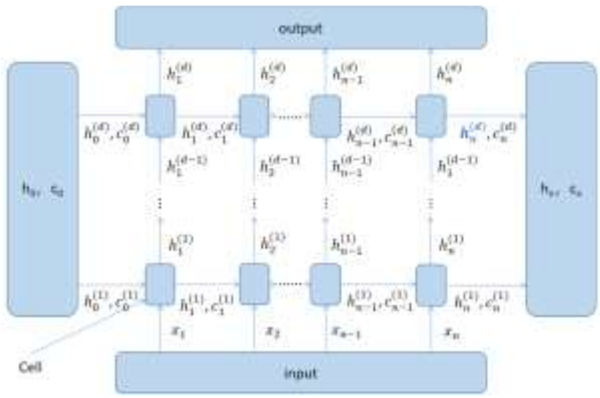

Figure.4 The framework of the model

First we need to construct the framework of the model, the overall framework of Convolutional Long Short-Term Memory is shown in Figure 3. Where h represents the hidden state, c represents the cell state, x represents the input, the subscript represents the step, the superscript represents the number of layers, the bottom rectangle represents the input, the top rectangle represents the output, the leftmost rectangle represents the initialization, the rightmost rectangle represents the final result of each layer, and the small rectangle in the middle part represents the cell module, which is a computational unit with inputs passed horizontally, hidden state, cell state and input images passed vertically, and outputs have new input images obtained by the computational formula. It is a computational unit, the inputs are the hidden state and cell state passed horizontally and the input image passed vertically, and the outputs are the new hidden state and cell state obtained by the computational formula, in which the hidden state should be passed horizontally and vertically.

After determining the model structure, the detection model can be trained using the dataset. In this paper, the model is trained using k-fold cross-validation (k=5) and the network parameters are tuned using the Adam optimizer. Since the training set has sufficient images and obvious features, only 100 epochs are needed for training. First, during the training process, after each batch is trained, clear the gradient of the optimizer and input the data into the model to get the model output, calculate the MSE loss and cumulative loss of the labels and outputs, and then back-propagate the MSE loss and update the optimizer, and finally, at the completion of training in one epoch, calculate the average loss based on the cumulative loss as the loss for the current round of training. other batches of data training and so on until the training is completed.

4.2 Model Testing

After the model has been trained it has to be tested as a way to determine how good the model is. This test is carried out using a previously established test set, the test method is to input eight consecutive images to the model, the model outputs the corresponding results, and this result is calculated with the labeling metrics to obtain the values of PSNR, SSIM and LPIPS. After getting the values of the metrics for all the samples in the test set using this method, the average of all the metrics is calculated to get the accuracy of the model. The real labeling is shown in Figure 5 and the output of the model is shown in Figure 6.

By calculation, the PSNR has a value of 40.44, SSIM has a value of 0.95, and LPIPS has a value of 0.14. From the previous description of these three metrics, we can get that when the value of PSNR is greater than 40, the output image of the model is of higher quality. When SSIM is closer to 1, the two images are more similar. When the value of LPIPS is closer to 0, it means that the two images are less and less different. The calculated values of all three metrics demonstrate the feasibility of the model for the application of detecting the trajectories of labeled particles.

Figure.5 True label

Figure.6 Output of the model

International Journal of Science and Engineering Applications
Volume 13-Issue 12, 28 – 31, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1312.1007

5. CONCLUSION

The purpose of this article is to explore the feasibility of Convolutional Long Short-Term Memory in other application areas. Previously Convolutional Long Short-Term Memory is mainly applied in the field of video prediction, after this research, the model shows its strength in detecting trajectories.

The feasibility of the model in synthesizing trajectories was successfully demonstrated by three metrics: PSNR, SSIM, and LPIPS. The method is not only simpler than the traditional image detection method, but also very convenient. No complex operations are required to use this model. Therefore, the application of Convolutional Long Short-Term Memory in the detection of labeled particles is also feasible.

6. REFERENCES

[1] MARIGO M, DAVIES M, LEADBEATER T, et al. Application of Positron Emission Particle Tracking (PEPT) to validate a Discrete Element Method (DEM) model of granular flow and mixing in the Turbula mixer [J]. International Journal of Pharmaceutics, 2013, 446(1): 46-58.

[2] ROMANENKO K, AVDIEVICH N. Unilateral RF sensors based on parallel-plate architecture for improved surface-scan MRI analysis of commercial pouch cells [J]. Journal of Magnetic Resonance Open, 2024, 18(100141.

[3] YAMAMOTO S, YOSHINO M, NAKANISHI K, et al. A comparative study of EM-CCD and CMOS cameras for particle ion trajectory imaging [J]. Applied Radiation and Isotopes, 2024, 204(111143.

[4] KORHONEN J, YOU J. Peak signal-to-noise ratio revisited: Is simple beautiful?; proceedings of the 2012 Fourth International Workshop on Quality of Multimedia Experience, F 5-7 July 2012, 2012 [C].

[5] ZHANG R, ISOLA P, EFROS A A, et al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric; proceedings of the 31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018, June 18, 2018 - June 22, 2018, Salt Lake City, UT, United states, F, 2018 [C]. IEEE Computer Society.

[6] LIU X Y, SPECHT E, MELLMANN J. Experimental study of the lower and upper angles of repose of granular materials in rotating drums [J]. Powder Technology, 2005, 154(2): 125-31.