

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

BMDSIS Project title:

Detecting and predicting behaviors of neurodegenerative disease primate models

LUYIYI [Signature]
Keng-Hwee Chiam [Signature]

SCHOOL OF BIOLOGICAL SCIENCES

2025

Introduction

1. Research Background

Neurodegenerative diseases [1] are common disorders resulting from the selective and progressive loss of neurons. The threat of such type of diseases is growing in recent years, and there has been a gradual increase in research as well as emphasis on the mechanisms of the disease. Richard Armstrong et al. [2] have suggested that some ‘prion’-like behaviors in the brain due to external environmental and internal genetic risk factors are responsible for the emergence of such diseases. Pierrick Coupé et al. [3] have recently analyzed and trained brain MRIs to construct a 'Lifespan Tree' to study the mechanisms and processes of different types of neurodegenerative disorders, such studies based on brain tissues are of crucial importance for disease generation and prevention.

The prediction of neurodegenerative diseases is equally important in the field of medicine, and recent research is currently focusing in another area on several areas of speech signal analysis, voice characterization [4], or movement characterization [5] related to fine joints (hands, eyes [6], etc.) analysis of patients, due to the fact that muscle abnormalities that can be fed back to movements can be the diagnostic criteria for neurodegenerative diseases [7]. With the advancement of machine learning, there is a proliferation of technological developments in deep learning, CV (CV, computer vision), and LLM (LLM, large language model) [8], which offer new possibilities for the prediction of neurodegenerative diseases.

Although significant progress has been made in research on prediction of neurodegenerative diseases based on physical sensor information and motion characterization, physical information acquisition has always suffered from poor convenience and implementation, as well as time response [9]. which is a drawback in the application of practical disease diagnosis. Since CV-based video pose estimation for neurodegenerative disease prediction is gradually being explored, medical practitioners as well as engineers are constantly looking for some easier and promising methods for disease risk prediction, for instance, a directly end-to-end video analysis with pose estimation [10].

2. Problem Statement

With the development of technologies such as CV and deep learning, prediction methods based on video analytics and posture estimation have gradually gained attention, but there are still many challenges for early diagnosis and prediction of neurodegenerative diseases (e.g., Parkinson's, Acromegaly, etc.). For example, when dealing behavioral data from animal models, video analytics methods have problems such as inaccurate processing of occlusion features (because of the uncertainty of animal movement, they are often caged [11]), as well as inadequate data preprocessing methods (there is no end-to-end data preprocessing solution exist now), which constrain the effectiveness of the practical application of video analytics in the prediction of neurodegenerative diseases [12]).

Therefore, how to develop a more effective data preprocessing method and video analytics procefure, or even ways to evaluate their optimization effects on the detection and prediction accuracy of animal behavior which lead to risk of neurodegenerative diseases is an urgent problem to be solved at present.

3. Research Aim and Innovation

In conclusion, this study aims to address this two following two issues:

1. Designing a preprocessing methods for video datasets of occluded animals in the hope of improving the estimation accuracy on prediction of the joint point positions as well as postures of moving animals.
2. Designing a video data preprocessing peocedure for the prediction of neurodegenerative diseases risk, which contributes to the future research on the improvement of prediction accuracy.

Besides, this study has the following innovations:

1. Use an unsupervised (CANNY filter and Closed Operatio) and a supervised method (AOT-GAN framework) to realize the preposseing for the video of caged moving animals.
2. Two methods are designed to quantify and visualize the accuracy of posture estimation of the caged marmoset dataset in terms of depth estimation results and distance evaluation metrics from the macro and micro perspectives, respectively.
3. The optimization of unsupervised versus supervised learning for cage-obscured animal video preprocessing schemes for primate pose estimation accuracy enhancement is verified using a new evaluation metric.
4. Designed an automated full process for filtering and selecting video datasets of caged marmosets, and verified the effectiveness of the method for improving the accuracy of primate pose estimation.

Methodology and Experiment\

1. Dataset Description

The dataset for this experiment was of six marmosets induced by MPTP, containing both female and male sexes. Marmosets were chosen because they (1) are genetically similar to humans to better mimic the human brain (2) have similar behaviors to humans (moving more slowly) and (3) are sensitive to MPTP, which has implications for the study of mechanisms of action of neurodegenerative disorders related to the brain. Secondly,: Studies have shown that MPTP-induced marmoset models can determine drug efficacy thresholds and thus help predict the success of clinical trials [13].



Fig 1. Pattern of the dataset

The capture range is shown (Fig. 2) as circle A is the range of motion of the marmoset, and the dotted lines a and b are the camera viewpoints of the marmoset (Fig. 3) Geometry A is the range of motion of the marmoset, and the dotted lines indicate the units and format of the coordinates. For each chamois marmoset dataset length I

intercepted the entire dataset containing up to a 30-second segment of the animal's locomotor state in order to minimize model processing time and to facilitate post-experimentation.

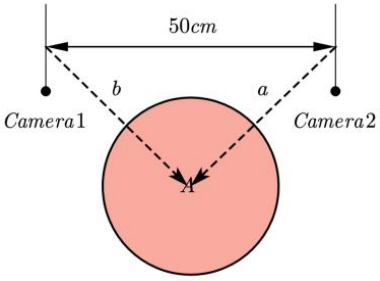


Fig 2. 2D-capture schema

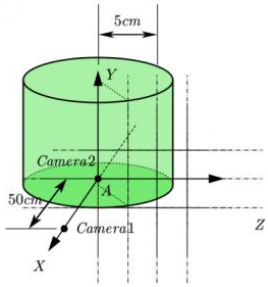


Fig 3. 3D-capture schema and the axis schema

In addition to this, I frame-pumped the video clips of each pom-pom marmoset according to the parameter interval=3 to compose the video-image dataset commonly used in the machine vision learning process, specifically, the size of the video and image datasets are as follows.

Tab 1. Detailed information of dataset

Name	If MPTP	Size of video recorded (unit:KB)	Number of picture	Size of picture (unit:MB)
Tutukueh	No	493857	270	638
Longtong	No	493897	270	647
Koi	Yes	493929	270	618
Horlick	No	494195	270	636
Briyani	No	493946	270	663
Boba	Yes	493911	270	625

2. Methodology

In summary, for the cut video dataset, I am using both supervised and supervised methods to filter the occlusion cages, and then the filtered video frame dataset is cut into regions of interest to reduce the computational resources, and then posture estimation is carried out, and the similarity of the posture estimation evaluation metrics is used to make judgments.The backtracking to the filtering step to filter the cages automatically filter the key frames which are useful for improving the prediction accuracy and retaining them only. help key frames and keep them only, and finally judge the actual attitude estimation optimization effect of the method using both macro and micro metrics. The overview diagram of the method in this paper is shown in the following figure.

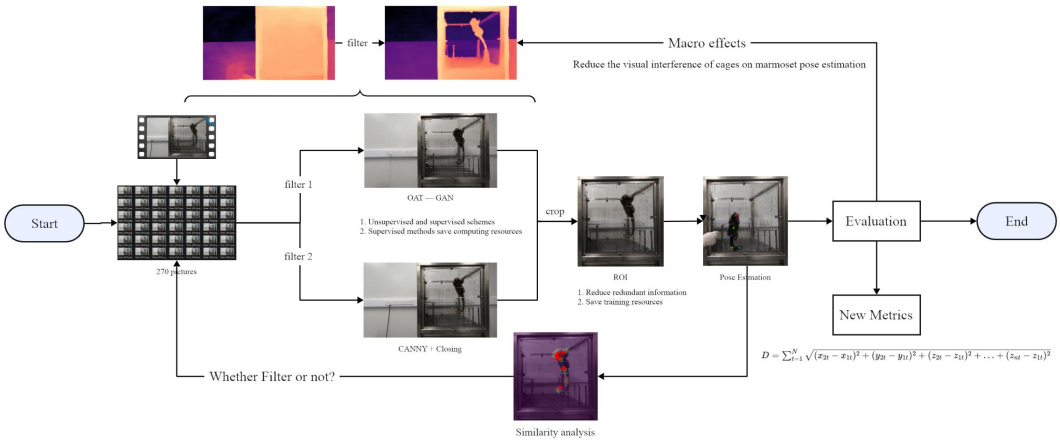


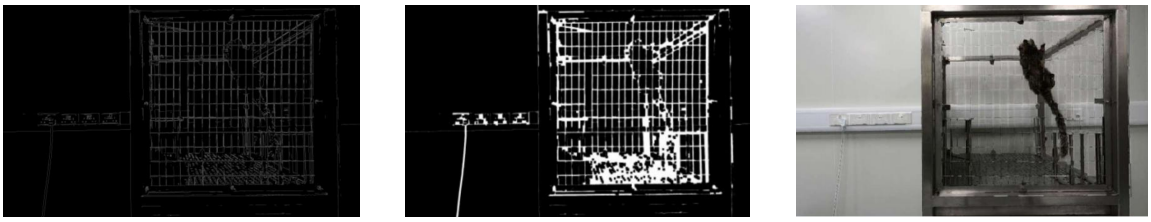
Fig 4. Abstract of the research method

3 Experimental Design

3,1 Filter design

(1) Unsupervised learning filtering method

Since the cage and the marmoset have very distinct linear features in the actual keyframe image, the Canny edge detector is first used to identify the object contours (as shown in Fig. 5(a)), then the contour interval is filled with a closed operation to get a complete region tiled image (as shown in Fig. 4(b), the cage features are clearer at this time), and then the cage portion is filtered by utilizing the mean padding on it. Fig. 5(c) illustrates the effect of the filtering, if a pixel is identified as a cage on the image, its pixel color will be replaced by the mean value of the colors of the 8 surrounding nearest neighbor pixels.

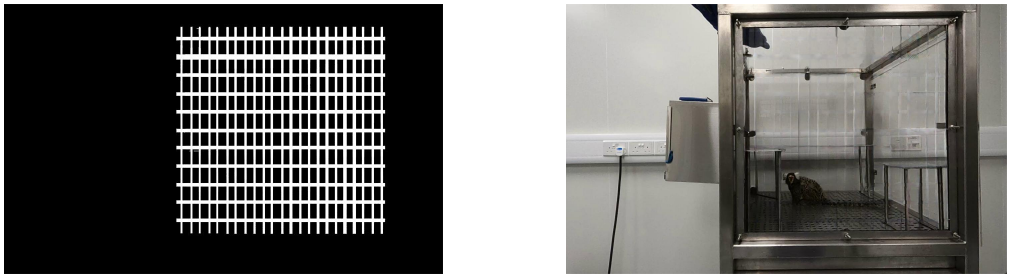


(a) (b) (c)

Fig 5. (a) Outcome of Canny detection (b) Outcome of Complete Method (c) Outcome of Cage filter

（2）*Supervised Learning Filtering Methods*

AOT-GAN (AOT-GAN, Aggregated Contextual Transformations for High-Resolution Image Inpainting) [14] enhances contextual inference by aggregating contextual transformations in the sensory field and a specific mask prediction task to train the discriminator, and is a model for Generative Adversarial Network Modeling for High-Resolution Image Restoration. While performing cage filtering on caged velvet monkeys, I attempted to fill and repair the image content of the cage portion by using the image repair function of AOT-GAN to generate an image without cage masking.



(a) (b)

Fig 6. (a)Hand-made mask for AOT-GAN (b) Filter outcome of AOT-GAN network

3.2 Region of Interest Selection

In order to reduce the memory cost and training time of the model training, after filtering the original sampled data, I performed a scaling of the dataset and selection of the region of interest, which can reduce the size of the dataset into the training iteration to save time cost.



(a) (b)

Fig 7. (a)Detection outcome for ROI (b) Cropped outcome of ROI

3.3 Model Training

DeepLabCut [15] is an open-source markerless pose estimation tool for animal behavior analysis, with multiple deep learning algorithms provided in the software to train neural networks that support the tracking of a wide range of animals and body parts that can be used for pose estimation. The study used this platform, set 7 keypoints (head, body, tailstart, tailmiddle, tailfinished, leftfoot, righhtfoot), chose the Resnet-50 framework [16] , set the training parameters of save_iter=500, max_iter=3000, and limited the input image size to 128x128 pixels.

3.4 Model Evaluation

In this experiment, two model evaluation methods were designed to present the model results, macroscopically using the 3d depth estimation results [17]), and microscopically creating a new distance evaluation metric using sum of Euclidean distances between predicted and actual points at all time points to describe the prediction effect.The mathematical expression for this evaluation indicator is as follows:

$$D = \sum_{t=1}^{N} \sqrt{(x_{1t+1} - x_{1t})^2 + (y_{1t+1} - y_{1t})^2 + (x_{2t+1} - x_{2t})^2 + \ldots + (y_{7t+1} - y_{7t})^2}$$

4 Results and Analysis

4.1 Comparison of Two Filter Method

The AOT-GAN retains more details of the monkeys, another is it significantly reduced preprocessing time, from several minutes to a few seconds for a single image. As estimated, unsupervised way of filter will take 4-5 minutes for one picture, while method of AOT-GAN will just take 4-5 seconds instead.



(a) (b)

Fig 8. (a)Picture filtered with unsupervosed method (b)Picture filtered with AOT-GAN

4.2 Macro effects of the method

Related studies have pointed out that the depth estimation results can be utilized to react to the accuracy of the animal's posture estimation, and therefore such logic is used in this paper. Fig.9(a) shows the results of animal posture estimation without cage estimation, and Fig.9(b) is the effect after processing. We can clearly find that when the cage is filtered before, the position and features of the velvet monkey are not accurately recognized by the computer algorithm, but when the cage is preprocessed using the method in this paper, the position and features of the velvet monkey are very clear. It shows that the effect is very valuable for improving the posture estimation results of the pom-pom monkeys and applying the results to the prediction of neurodegenerative diseases.



(a)



(b)

Fig 9. (a) Depth Estimation outcome before filtering (b) Depth Estimation outcome after filtering

4.3 Micro effects of the method

Before and after filtering, all six monkeys improved their predictions overall, with the most significant improvement being Boba, where the error rate was reduced to 50% of the original error, suggesting that the method was effective in improving the results of the monkey's posture estimation. However, the method performed poorly on the dataset on Horlick, and we need to analyze the reasons for this situation as well.

Tab 2. Filter outcome on six marmosets

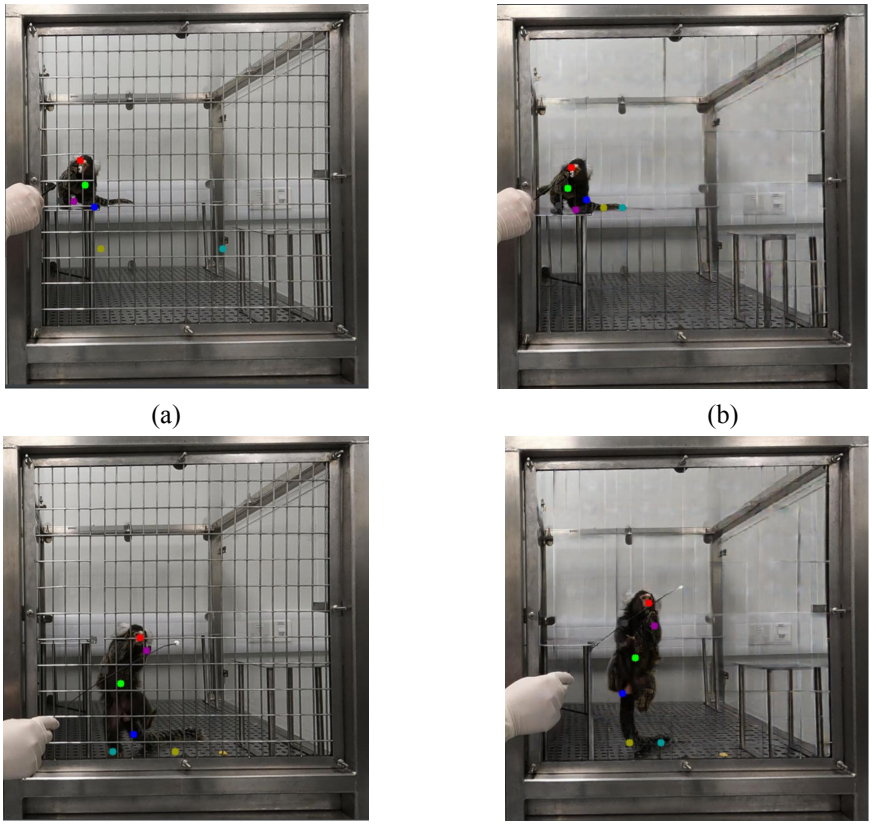
Name	head	body	tailstart	tailmiddle	tailfinished	leftfoot	righhtfoot	all
Boba	**33.68%**	**41.75%**	**64.86%**	**55.22%**	**47.07%**	**36.26%**	**55.53%**	**48.64%**
Briyani	-6.44%	**9.34%**	**25.22%**	**26.62%**	**68.45%**	**30.72%**	-85.18%	**23.28%**
Horlick	-38.64%	-55.11%	**28.49%**	-19.70%	-94.04%	-121.20%	-119.31%	-55.66%
Koi	**15.85%**	**22.35%**	-45.71%	-62.40%	-28.36%	-9.01%	-44.19%	-17.66%
Lontong	-23.00%	**25.57%**	**7.74%**	**46.16%**	-2.25%	**54.92%**	**48.59%**	**25.78%**
Tutukueh	-46.04%	**39.34%**	**50.17%**	**25.95%**	**20.00%**	**9.13%**	-7.53%	**14.11%**

If we focus only on just one marmosetperformance (take Briyani as an example) on the entire time series, all five parts of this one monkey's accuracy are improved before and after filtering, which shows that the method is effective in improving the prediction progress of small joints again.

Tab 2. Error distance of Briyani before and after filter

Index	Error	head	body	tail1	tail2	tail3	leftfeet	rightfeet
original	Average	6.70	5.81	10.91	11.47	15.70	8.10	5.04
	Sum	133.98	116.21	218.12	229.49	313.94	161.97	100.73
filtered	Average	7.13	5.27	8.16	6.50	4.95	5.61	9.33
	Sum	142.62	105.36	163.10	130.00	99.06	112.21	186.53
optimize	Average	-6.44	**9.34**	**25.22**	**43.35**	**68.45**	**30.72**	-85.18
	Sum	-6.44	**9.34**	**25.22**	**43.35**	**68.45**	**30.72**	-85.18

This is also illustrated by the visualization results in the figure below, which shows that Briyani's and Horlick's caudal features (blue dots), and left and right foot features (dark blue and purple dots) perform more accurately after filtering in (b), but without filtering in (a), the prediction of these key points produces a shift in the results, and such a posture estimation result is clearly detrimental to the prediction of Parkinson's.The same outcome also shown at (c) and (d) in Fig.10.



(a) (b)

(c) (d)

Fig 10. (a) Depth Estimation before filtering(Briyani) (b) Depth Estimation after filtering(Briyani) (c) Depth Estimation before filtering(Horlick) (d) Depth Estimation after filtering(Horlick)

If we only focus on the optimization of Boba's all time step injury, we can also get a very good result from the line graph of the results. Figure 11 & 12 shows the distribution of the optimization rate of Boba's 7 key points at all time points, and we can find that the accuracy of Boba's attitude estimation has been improved at almost all time points, which also shows the effectiveness of the method.

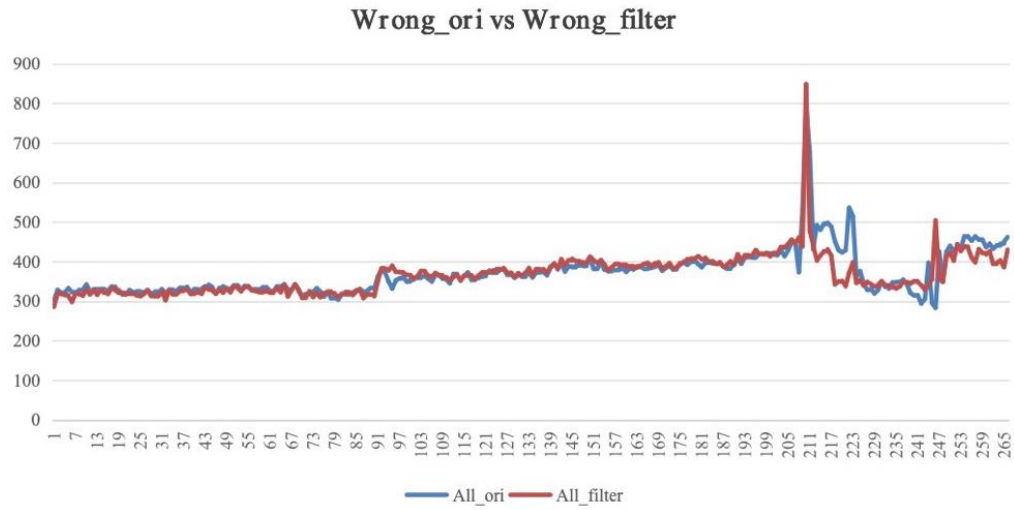


Fig 11. Error of pose estimation of Boba on all time steps

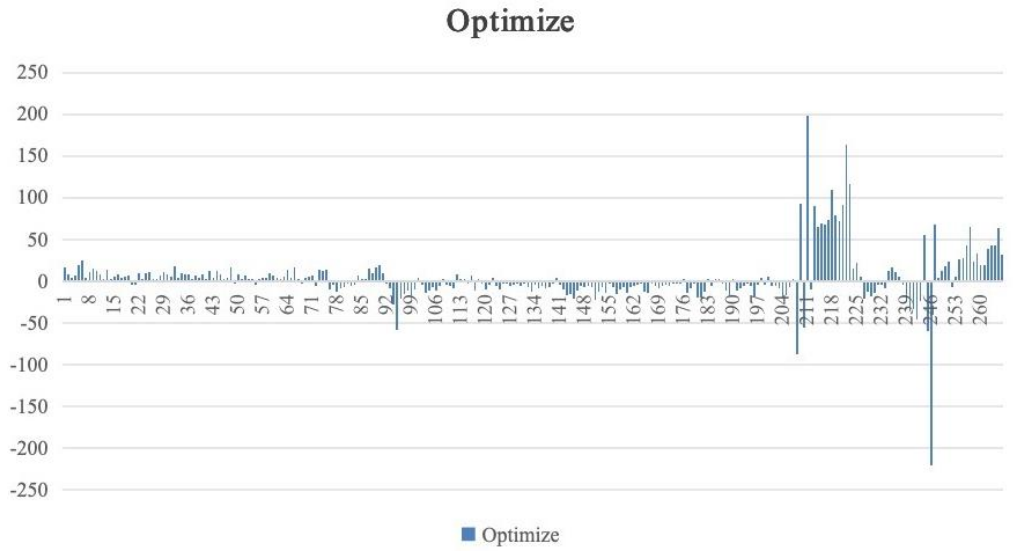
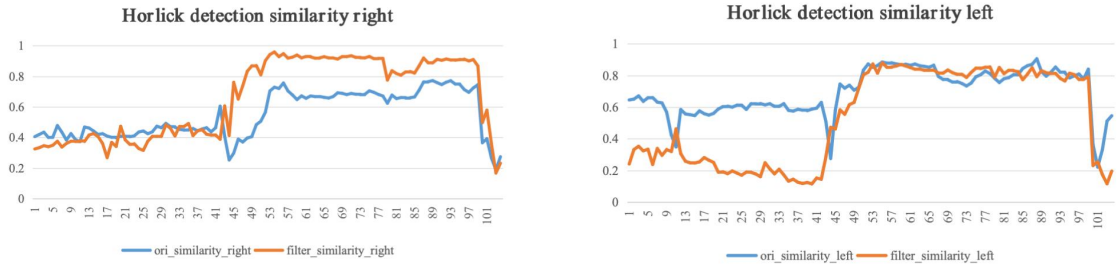


Fig 12. Optimize of pose estimation of Boba on all time steps

Theoretically, if the cages are filtered, the more obvious information will make the joint features more distinctive, but for horlick in the first few seconds the filtering operation rather reduces the accuracy of the recognition.I also tried some methods to find the cause of this problem, by analyzing the similarity of all time points, I found that horlick in the first few time points of the prediction.



(a) (b)

Fig 12. (a) Estimation similarity of Horlick (original) (b) Estimation similarity of Horlick (filtered)

By analyzing the original video, I found that this is because in the first few key frames, the monkey's hand is in contact with the cage's border, which leads to the filtering process of the cage, which also filters the hand's key position information. Therefore, we can use the confidence level of the results to go back to the very beginning of the filtering process, and automate the process of selecting which points in time are not suitable for utilizing the filtering method, and those points in time where the method can be applied, which also provides a guide for automated video data preprocessing in later stages in the medical field.

Conclusion

In this paper, two unsupervised and supervised data preprocessing methods for mask occlusion feature animal medical video data are proposed to improve the prediction accuracy of pose estimation animal medical dataset. The effectiveness of the two filtering methods in optimizing the accuracy of pose estimation is affirmed, and the results show that the AOT-GAN method has a significant advantage in detail preservation and preprocessing time, and is able to complete the processing of a single image in a few seconds, whereas the traditional unsupervised method based on Canny with closure operations takes several minutes.

By comparing the depth estimation results and keypoint prediction errors before and after filtering, we find that the filtered data can more accurately identify the position and features of the velvet monkeys, thus improving the accuracy of the pose estimation, and two thirds of the six velvet monkeys analyzed **were optimized with 20-60% pose estimation accuracy, which was optimized throughout the entire time series.** After analyzing the posture estimation time points that produced bad results, I also concluded a processing logic that can be used to automate video preprocessing using the confidence before and after preprocessing, as shown below. This scheme provides a new technical means for video prediction of neurodegenerative diseases, and has a very high technical potential in the practical application of video analysis in animal medicine, which can achieve more accurate high-precision pose estimation, better understanding of animal behavior and disease progression, and support for early diagnosis and intervention of diseases.

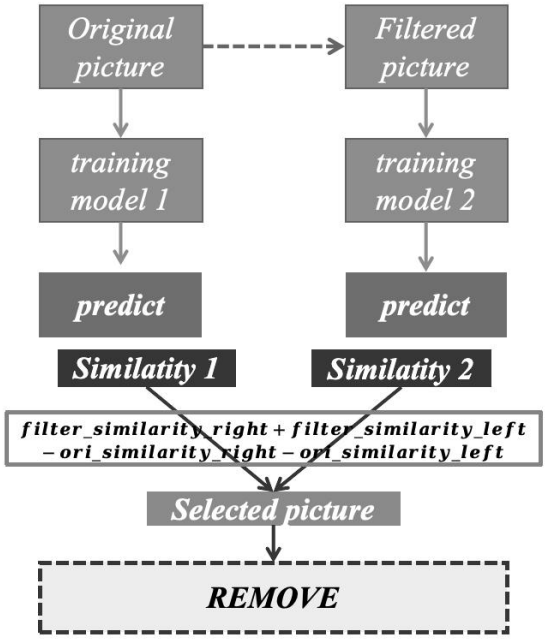


Fig 13. A logic could be used in auto preposessing framework based on similarity of prediction

Future research directions will focus on further optimizing the data preprocessing methods to improve the adaptability to different types of animal models and disease states. In addition, we will explore how these techniques can be applied to broader studies of neurodegenerative diseases, such as Alzheimer's disease and amyotrophic lateral sclerosis (ALS), with a view to providing new ideas and tools for the early detection and treatment of these diseases. We will also continue to improve the automation of the model so that it can more intelligently recognize and process key frames in the video data, thereby further improving the accuracy and efficiency of the predictions.

Reference

[1] Dugger, Brittany N., and Dennis W. Dickson. "Pathology of neurodegenerative diseases." Cold Spring Harbor perspectives in biology 9.7 (2017): a028035.

[2] Armstrong, Richard. "What causes neurodegenerative disease?." Folia neuropathologica 58.2 (2020): 93-112.

[3] Coupé, Pierrick, et al. "Lifespan tree of brain anatomy: diagnostic values for motor and cognitive neurodegenerative diseases." arXiv preprint arXiv:2502.09682 (2025).

[4] Çelik, Burak, and Ayhan Akbal. "Distinguishing Parkinson's Patients Using Voice-Based Feature Extraction and Classification." arXiv preprint arXiv:2501.14390 (2025).

[5] Matsumoto, Masahiro, et al. "Machine Learning-Based Differential Diagnosis of Parkinson's Disease Using Kinematic Feature Extraction and Selection." arXiv preprint arXiv:2501.02014 (2025).

[6] Salazar, Isail, et al. "A convolutional oculomotor representation to model parkinsonian fixational patterns from magnified videos." Pattern Analysis and Applications 24.2 (2021): 445-457.

[7] Uitti, Ryan J., et al. "Neurodegenerative 'overlap' syndrome: clinical and pathological features of Parkinson's disease, motor neuron disease, and Alzheimer's disease." Parkinsonism & related disorders 1.1 (1995): 21-34.

[8] Jantre, Sanket, et al. "Uncertainty-Aware Adaptation of Large Language Models for Protein-Protein Interaction Analysis." arXiv preprint arXiv:2502.06173 (2025).

[9] Jing Y, Qin P, Fan X, Qiang W, Wencheng Z, Sun W, Tian F, Wang D. Deep Learning-Assisted Gait Parameter Assessment for Neurodegenerative Diseases: Model Development and Validation. J Med Internet Res. 2023 Jul 5;25:e46427. doi: 10.2196/46427. PMID: 37405831; PMCID: PMC10357315.

[10] Yee, David Vun Sing, Robert Bowers, and Anthony McGarry. "Vision-based motion capture for the gait analysis of neurodegenerative diseases: A review." Gait & Posture (2024).

[11] Fang, Cheng, et al. "The posture detection method of caged chickens based on computer vision." Animals 14.21 (2024): 3059.

[12] Sun, Zheng, et al. "Caged marmoset Dataset: A New Benchmark for Caged marmoset Pose Estimation." Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Cham: Springer Nature Switzerland, 2022.

[13] Beaudry, F., Huot, P. The MPTP-lesioned marmoset model of Parkinson's disease: proposed efficacy thresholds that may potentially predict successful clinical trial results. J Neural Transm 127, 1343–1358 (2020). https://doi.org/10.1007/s00702-020-02247-2.

[14] Zeng, Yanhong, et al. "Aggregated contextual transformations for high-resolution image inpainting." IEEE transactions on visualization and computer graphics 29.7 (2022): 3266-3280.

[15] Mathis, Alexander, et al. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning." Nature neuroscience 21.9 (2018): 1281-1289.

[16] Koonce, Brett, and Brett Koonce. "ResNet 50." Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization (2021): 63-72.

[17] Barnard, Shanis, et al. "Quick, accurate, smart: 3D computer vision technology helps assessing confined animals' behaviour." PloS one 11.7 (2016): e0158748.