UNIVERSITY OF TURKU

Facial Paralysis Grading Based on Dynamic and Static Features

Master's thesis (Technology)
Master's Degree Programme in Information and Communication Technology
Department of Computing, Faculty of Technology
Master of Science in Technology Thesis

Author:
Bolu Wang

Supervisors:
Dr. Paavo Nevalainen
Prof. Tomi Westerlund

June 2023

Pro gradu -tutkielma
Tietotekniikan laitos, Teknillinen tiedekunta
Turun yliopisto

TUTKIJA: BOLU WANG
Otsikko: Facial Paralysis Grading Based on Dynamic and Static Features
Tutkinto-ohjelma: Tieto- ja viestintätekniikka
Sivumäärä: 51 sivua
Päivämäärä: kesäkuu 2023

Perifeerinen kasvojen hermohalvaus, joka tunnetaan myös nimellä kasvojen halvaus (FP), on yleinen kliininen sairaus, joka vaatii subjektiivista arviointia ja FP -asteikon pisteytystä. Joitakin automaattisia kasvohalvauksen luokittelumenetelmiä on olemassa, mutta yleensä ottaen ne punnitsevat vain joko staattisia tai dynaamisia piirteitä. Tässä tutkielmassa ehdotetaan automaattista kasvojen halvaantumisen arviointimenetelmää, joka kattaa sekä staattiset että dynaamiset ominaisuudet.

Menetelmän ensimmäinen vaihe suorittaa ensin esikäsittelyn kohteiden kerätyille kasvojen ilmevideoille, mukaan lukien karkea videon sieppaus, videon vakautus, avainruudun poiminta, kuvan geometrinen normalisointi ja harmaasävyjen normalisointi. Seuraavaksi menetelmä valitsee avainruuduiksi ilmeettömän tilan ja kasvojen ilmeiden maksimitilan kuvadatasta kerryttäen tutkimuksen data-aineiston.

Tietojen esikäsittely vähentää virheitä, kohinaa, redundanssia ja jopa virheitä alkuperäisestä datasta. Kuvan staattisten ja dynaamisten piirteiden poimimisen perusta on käyttää Ensemble of Regression Trees -algoritmia 68 kasvojen merkkipisteiden määrittämiseen. Merkkipisteiden perusteella määritellään kuvan kiinnostavat alueet. Horn-Schunckin optisen virtausmenetelmän mukaisesti poimitaan optisen virtauksen tiedot joistakin kasvojen osista, ja dynaaminen luonnehdinta lasketaan vasempien ja oikeiden osien välille.

Lopuksi dynaamisen ja staattisen piirteiden luokittelun tulokset painotetaan ja analysoidaan kattavasti koehenkilöiden FP-luokitusten saamiseksi. 32- ulotteinen staattisten piirteiden vektori syötetään tukivektorikoneeseen luokittelua varten. 60-ulotteinen dynaamisten piirteiden ominaisuusvektori syötetään pitkän ja lyhyen aikavälin muistiverkkoon luokittelua varten. Parhaan luokittelijan tarkkuus, täsmällisyys, palautustaso ja f1 saavuttavat arvot 93,33%, 94,29%, 91,33% ja 91,87%.

Asiasanat: kasvojen halvaus, kasvojen maamerkit, optinen virtaus, koneoppiminen, kattava analyysi

Master of Science in Technology Thesis
Department of Computing, Faculty of Technology
University of Turku

RESEARCHER: BOLU WANG
Title: Facial Paralysis Grading Based on Dynamic and Static Features
Programme: Master's Degree Programme in Information and Communication Technology
Number of pages: 51 pages
Date: June 2023

Peripheral facial nerve palsy, also known as facial paralysis (FP), is a common clinical disease, which requires subjective judgment and scoring based on the FP scale. There exists some automatic facial paralysis grading methods, but the current methods mostly only consider either static or dynamic features, resulting in a low accuracy rate of FP grading. This thesis proposes an automatic facial paralysis assessment method including both static and dynamic characteristics.

The first step of the method performs preprocessing on the collected facial expression videos of the subjects, including rough video interception, video stabilization, keyframe extraction, image geometric normalization and gray-scale normalization. Next, the method selects as keyframes no facial expression state and maximum facial expression state in the image data to build the the research data set.

Data preprocessing reduces errors, noise, redundancy and even errors in the original data.
The basis for extracting static and dynamic features of an image is to use Ensemble of Regression Trees algorithm to determine 68 facial landmarks. Based on landmark points, image regions of image are formed. According to the Horn-Schunck optical flow method, the optical flow information of parts of the face are extracted, and the dynamic characteristics of the optical flow difference between the left and right parts are calculated.

Finally, the results of dynamic and static feature classification are weighted and analyzed to obtain FP ratings of subjects. A 32-dimensional static feature is fed into the support vector machine for classification. A 60-dimensional feature vector of dynamical aspects is fed into a long and short-term memory network for classification. Videos of 30 subjects are used to extract 1419 keyframes to test the algorithm. The accuracy, precision, recall and f1 of the best classifier reach 93.33%, 94.29%, 91.33% and 91.87%, respectively.

Keywords: facial paralysis, facial landmarks, optical flow, machine learning, comprehensive analysis

Table of Contents

Chapter I Introduction

1.1 Research Background

1.1.1 Research significance and purpose

Peripheral facial nerve palsy is a common clinical and frequently-occurring disease, with an annual incidence of about 15-40/100,000, and it is not restricted by age or gender[1]. It is often caused by viral infection, trauma, intracranial and extracranial tumors, inflammation of the pharynx or external auditory canal. It may also be caused by inflammation, ischemia or hemorrhage of the pons or medulla oblongata, but facial neuritis is the most common.

Sudden facial paralysis causes great harm to the patient's physiology and impacts the patient's psychological state. In the disease diagnosis and treatment process, some patients have interrupted treatment due to poor psychological conditions, and their prognoses are unsatisfactory due to the long-term treatment course and repetitive treatment measures. Therefore, the doctor should also strengthen psychological intervention for them while treating the physical symptoms of patients with facial paralysis, and informing the patients and their families of the cause, disease manifestations, disease course, and prognosis promptly[2]. Accurate recognition and evaluation of facial paralysis is the basis for effective treatment and can also provide an accurate reference for psychological intervention.

The current clinical facial paralysis evaluation method is that doctors require the patient to make various facial movements according to the scoring items of the facial paralysis evaluation scale. They subjectively observe the facial appearance and facial muscle movement state for classification. There are dozens of scales for clinical evaluation of facial nerve function, including House-Brackmann facial nerve grading system[3], Yanagihara grading system[4], The Nottingham System (NS) [5], Sunnybrook grading system[6], Degree of Facial Nerve Paralysis Scale (DFNP) [7], Facial Disability Index (FDI) [8], etc.

Li, Gao et al.[9] compared the evaluation effects of peripheral facial nerve palsy on the scales of House-Brackmann (1985), House-Brackmann 2010 and Sunnybrook (1996). It is believed that the classification of House-Brackmann (2010) and Sunnybrook (1996) scales can evaluate the function of the peripheral facial nerve

more accurately, and it is easier for doctors to grasp. The H-B scale is also one of the most widely studied scales in the world[10][11]. Therefore, this thesis chooses the H-B scale as the basis for facial paralysis, which is shown in Table 1.

Table 1 House-Brackmann scale standards[3]

Grade	Description	Characteristics
I	Normal	Normal facial function in all areas
II	Mild dysfunction	Gross: slight weakness noticeable on close inspection; may have very slight synkinesis At rest: normal symmetry and tone Motion: Forehead: moderate to good function Eye: complete closure with minimum effort Mouth: slight asymmetry
III	Moderate dysfunction	Gross: obvious but not disfiguring difference between two sides; noticeable but not severe synkinesis, contracture, and/or hemifacial spasm At rest: normal symmetry and tone Motion: Forehead: slight to moderate movement Eye: complete closure with effort Mouth: slightly weak with maximum effort
IV	Moderately severe dysfunction	Gross: obvious weakness and/or disfiguring asymmetry At rest: normal symmetry and tone Motion: Forehead: none Eye: incomplete closure Mouth: asymmetric with maximum effort
V	Severe dysfunction	Gross: obvious but not disfiguring difference between two sides; noticeable but not severe synkinesis, contracture, and/or hemifacial spasm At rest: asymmetry Motion: Forehead: none Eye: complete closure Mouth: slightly movement
VI	Total paralysis	No movement

To a certain extent, doctors rely on their own experience to evaluate facial paralysis with the scale. Therefore, doctors who do not have enough experience in diagnosis and treatment may misjudge the patient's condition. The prognostic diagnosis of facial paralysis requires re-scoring the patient's condition after treatment. The doctor makes diagnosis and treatment adjustments based on several previous evaluations of the treatment effect. The frequency of follow-up visits for facial paralysis, however, is often measured in weeks. Doctors need to see a large number of patients every day. Even if there are medical records, it is still difficult for patients

returning to the clinic to recall the details of previous consultations. At the same time, there is also the possibility that patients may consult multiple doctors. While computer-based automatic facial paralysis recognition and evaluation can establish a complete database for doctors and patients, assist doctors in more accurate diagnosis and treatment, and provide objective data preparation to optimize treatment plans. It can also allow patients to have an intuitive feeling about the effect of treatment, and make them psychologically prepared for prognosis.[12].

Computer-based automatic facial paralysis recognition and grading assessment combine information technology with medical diagnosis to assist doctors in accurately diagnosing and treating facial paralysis patients. The computer-based facial paralysis evaluation method only needs to obtain a video of the patient. In the video, the patient makes several expressions as required, such as no expression, raised eyebrows, closed eyes, bulged cheeks, and showing teeth. The initial data of the evaluation algorithm is these specific expression figures and basic expressionless figures. At present, there are many artificial intelligence facial paralysis recognition and evaluation methods, which can be roughly divided into two categories: facial image features manual extraction (Figure 1(a)) and neural network feature extraction (Figure 1(b)) [13]. The architecture of the evaluation system is shown in Figure 1.

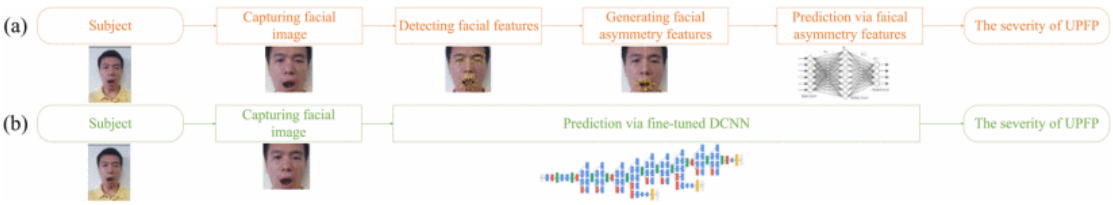


Fig.1 Computer-based facial paralysis evaluation method process [13]
(a) Traditional objective evaluation process
(b) Deep convolutional network evaluation process

1.2 Research Status of This Area

1.1.2 Facial Paralysis Evaluation by Manually Extracting Features of Patient Facial Images

The basic idea of the facial paralysis evaluation method that manually extracts patient facial image features is to establish the mapping relationship between the patient facial image and the grade of the facial paralysis scale. Considering that the diagnostic method based on the facial paralysis scale compares the symmetry and motor ability of the face, facial landmarks or regions of interest such as eyebrows,

eyes, nose, and mouth are closely related to symmetry. Therefore, the key issue in establishing the mapping relationship is the determination of facial landmarks and regions of interest, and the feature extraction on this basis.

Facial landmarks positioning method: The facial landmarks used to evaluate facial paralysis are mainly located on the contours of eyes, brows, nose and mouth. They can be confirmed through edge detection algorithms or machine learning algorithms.

The location method of the region of interest (ROI): compare the facial image with a normalized facial region distribution map (Figure 2), then the facial ROI can be located.

After determining the corresponding landmark point or ROI, we can calculate perimeters, area, height, angle, Gaussian curvature, coordinates, displacement and other parameters to obtain the characteristics[14].

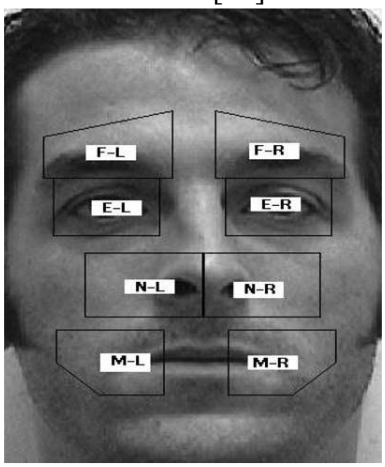


Fig.2. Region of Interest[16]

After extracting the features, I can screen out the features linearly related to the scale score as the input of the classifier. Train and test the classifier and obtain the predicted classification result at last. The following introduces several typical artificially selected features to evaluate patient facial images.

Guo et al.[14] proposed a non-invasive computer assessment framework for unilateral peripheral facial paralysis (UPFP). This method selects the H-B scale as the criterion, intercepts the facial area of the acquired initial image data (as shown in Figure 3), and applies the face alignment algorithm Supervised Descent Method (SDM) to obtain the landmark points of the face[15]. Then it performs feature screening and finally trains the classifier.

The idea of the algorithm: assume that I have initial mark points $\boldsymbol{x_0}$, and hope to find the accurate feature points $\boldsymbol{x}_*$ step by step through iteration. Given an image

containing m pixels $\boldsymbol{d} \in \boldsymbol{R^{m\times 1}}$, $\boldsymbol{d(x)} \in \boldsymbol{R^{p\times 1}}$ is used to index p feature points of the image, and x represents p feature points. $\boldsymbol{h}(\mathbf{d}(\boldsymbol{x})) \in \boldsymbol{R^{128p\times 1}}$ represents the SIFT (Scale Invariant Feature Transform) feature vector. SIFT was proposed by David Lowe in 1999 **Virhe. Viitteen lähdettä ei löytynyt.** and improved in 2004 **Virhe. Viitteen lähdettä ei löytynyt.**. It is a robust and effective method of image feature extraction that can adapt to different situations. The feature vector obtained by this algorithm is a 128-dimension vector. $f(\boldsymbol{x_0}+\boldsymbol{\Delta x})$ represents the difference between the current predicted feature point and the real feature point. At the minimum value, the optimal position $\mathbf{d}(\boldsymbol{x_0}+\boldsymbol{\Delta x})$ of p feature points is obtained.

$$f(\boldsymbol{x_0}+\boldsymbol{\Delta x}) = \|\mathbf{h}(\mathbf{d}(\boldsymbol{x_0}+\boldsymbol{\Delta x})) - \phi_*\|_2^2 \quad (1.1)$$

$\phi_* = \mathbf{h}(\mathbf{d}(\boldsymbol{x}_*))$ is the SIFT feature vector of p marker points manually calibrated。

$$f(\boldsymbol{x_0}+\boldsymbol{\Delta x}) \approx f(\boldsymbol{x_0}) + \boldsymbol{J}_f(\boldsymbol{x_0})^T\boldsymbol{\Delta x} + \frac{1}{2}\boldsymbol{\Delta x^T}\mathrm{J}(\boldsymbol{x_0})\boldsymbol{\Delta x} \quad (1.2)$$

Find the derivative of $\boldsymbol{\Delta x}$， Let the derivative $\boldsymbol{f}'(\boldsymbol{x_0}+\boldsymbol{\Delta x})$ equal to zero, we can get:

$$\Delta \boldsymbol{x}_1 = -\mathrm{H}^{-1}\mathrm{J}_f = -2\mathrm{H}^{-1}\mathrm{J}_h^T(\phi_0 - \phi_*),\ \phi_0 = \mathbf{h}\big(\mathbf{d}(\boldsymbol{x_0})\big) \qquad (1.3)$$

Let $\mathrm{R} = -2\mathrm{H}^{-1}\mathrm{J}_h^T$, $\Delta\phi = \phi_0 - \phi_*$, can be regarded as the gradient direction. To simplify the model, we can transform nonlinear problems into linear problems.

$$\Delta \mathrm{x}_1 = \mathbf{R}_0\phi_0 + \mathbf{b}_0 \qquad (1.4)$$

Use training samples to learn $(\mathbf{R}_0 + \mathbf{b}_0)$, and use iteration to complete the convergence of feature points x

$$\mathrm{x}_k = \mathrm{x}_{k-1} - 2\mathrm{H}^{-1}\mathrm{J}_h^T(\phi_{k-1} - \phi_*) \qquad (1.5)$$

$$\mathrm{x}_k = \mathrm{x}_{k-1} + \mathbf{R}_{k-1}\phi_{k-1} + \mathbf{b}_{k-1} \qquad (1.6)$$

after 4 to 5 iterations, the location of the feature points can be well determined.

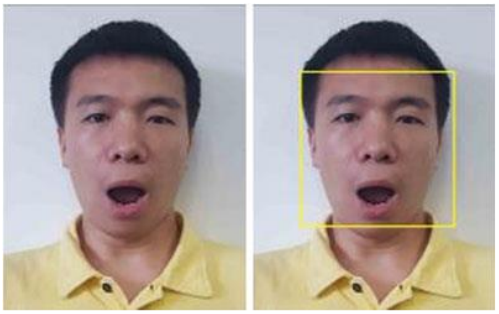


Fig.3. Extraction of the facial region[14]

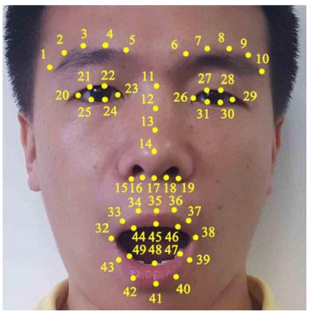


Fig.4. Determination of mark points[14]

SDM algorithm can get 49 facial landmark points (Figure 4). These landmark points can be divided into eight categories, as shown in Figure 5.

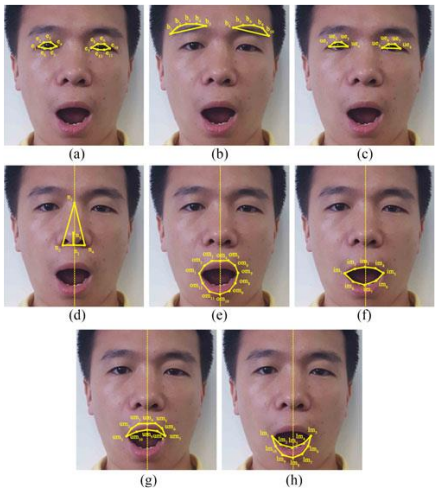


Fig.5. Groups of facial marking points

(a) eyes (b) eyebrow (c)upper eyelid (d)nose (e)outer profile of mouth (f)inner profile of mouth (g)upper lip (h)lower lip[14]

The features selected in the facial paralysis diagnosis and evaluation are usually associated with the shape symmetry, position symmetry, and displacement symmetry of the corresponding facial landmark points on both sides of the face. We can obtain the features by calculating the perimeter, area, height, angle, Gaussian curvature, coordinates, and displacement of the corresponding landmarks or regions of interest. By evaluating the correlation between the features and the linear regression of the H-B scale, they screen out the features whose R Square exceeds a certain threshold. And Support Vector Machine (SVM) is used to classify the patients into 6 categories.

This is what the experiment result tells us. The complete accuracy rate is 49.9%, the loose accuracy rate (the scoring error is within 1 level) is 87.97%, and the loose accuracy rate after sample argument is 90.01%.

This method selects static features as the input of the classifier with good robustness. Compared with the research which selects dynamic features, their amount of calculation is smaller. The limitation is that no time-related features are selected so that part of the facial asymmetry information is lost affecting the algorithm accuracy. The resolution of adjacent levels will be blurred. In addition, the automatic evaluation system designed by it is more complicated to operate and requires a certain learning period.

Ngo, Seo, Matsushiro et al.[16] proposed a facial paralysis assessment method based on Gabor features and Local Binary Pattern (LBP) features. This method also chooses the H-B scale as the criterion. It adjusts the faces according to the vertical line of the pupillary distance, then uses Adaboosting to obtain the area of interest of

the face, including the left and right eyebrows, left and right eyes, left and right nose, and left and right mouth. Convert each ROI into an LBP image, get the feature value after Gabor filtering, and use an SVM classifier to achieve facial paralysis classification.

The core concept of this method is to evaluate the degree of facial paralysis through the asymmetry of facial images, which improves the accuracy of facial paralysis to a certain extent. But it takes a long time to obtain the rating results because of the long response time of the Gabor filter. Its parameter selection is relatively simple and cannot fully describe the characteristics of facial paralysis, which makes the complete accuracy rate of facial paralysis rating not high. The final complete accuracy rate reaches 60.7%, and the loose accuracy rate is 91.3%.

Ridha and Shehieb et al.[19] designed a device for assisting facial paralysis rehabilitation in 2020. The software part running on the mobile phone uses the Google Machine Learning Kit to obtain facial landmarks (Figure 6); and then uses the standardized difference between pupils to the corner of the mouth on both sides (Figure 7). The network is trained to obtain the discrimination threshold. Then a standardized threshold is selected as the criterion for distinguishing facial paralysis. They select no scale as the criterion, and there is no explanation of the accuracy of the network judgment.

The advantage of this method is that the amount of calculation is very small. Only one feature is chosen for judgment so that the result is fast, suitable for running on mobile devices with low computing power. The disadvantage is that the judgment is not based on the clinical evaluation scale, and the results cannot be directly used as the basis for clinical diagnosis. In the case of small samples, whether the neural network algorithm model is accurate or over-fitting needs to be further evaluated.

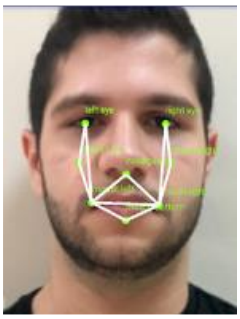


Fig.6. determination of facial mark points[19]

Fig.7. Feature selection result[19]

The commonality of the above three artificial feature extraction methods is to extract symmetry-related static features based on the landmarks or regions of interest

(including eyebrows, eyes, nose and mouth).

Among them, the method [14] is relatively professional, well robust, and applicable to auxiliary diagnosis, but there is still room for improvement on accuracy, and further research is needed. The method[16] is from early research with strong professionalism but slower reaction time and lower accuracy. The advantage of the method **Virhe. Viitteen lähdettä ei löytynyt.** is its convenience and is suitable for situations that require rapid detection and movement detection. However, the concept is very reasonable. Whether the feature selection is optimal and the accuracy threshold determination result remains to be discussed.

1.1.3 Evaluation of facial paralysis based on neural network and facial images of patients

To solve the problem of the low accuracy of the current artificial facial paralysis feature extraction method, many automatic facial paralysis assessment researches turn their attention to deep neural networks.

Hsu and Chang[20] proposed a deep hybrid network to realize the automatic quantitative analysis of facial paralysis. The idea is to extract facial contours and landmark points through a neural network, and then use the network to distinguish facial paralysis on the image.

This method constructs a Deep Hybrid Network (DHN) composed of 3 networks. The specific workflow is that the data image passes through the first network Netf segmenting raw image to get the face image. The face image is then input into the second network Netm to obtain a network of contour line segments connected by adjacent marking points (Line Segment Network) and a network to further obtain the marking points of the face image （Double Dropout Network）. Finally, the outputs of Netf and Netm are used as the input of the third-level network to detect and label the facial paralysis area (Figure 8). This method only detects and labels facial paralysis.

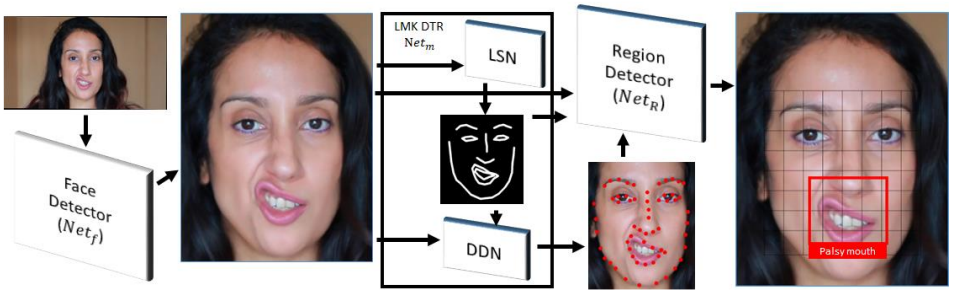


Fig.8. Deep hybrid network for automatic quantitative analysis of facial paralysis[20]

This method introduces a new three-segment network structure. Hsu utilize network learning to determine the location of facial paralysis and the degree of facial paralysis, without extracting features from landmarks or regions of interest. However, the study did not use the evaluation criteria of samples and scales marked by professional doctors, and the grading results obtained were not very authoritative. It remains to be further studied whether its conclusions can be used in clinical practice.

Guo et al.[13] proposed a deep convolutional neural network to automatically assess UPFP. This method selects the H-B scale as the criterion. It uses GoogLeNet to obtain the nonlinear mapping relationship between the patient facial image and the H-B grading. This method has fine-tuned the number of samples and data expansion and optimized the parameters of the convolutional network. The accuracy rate of facial paralysis evaluation reaches 91.25%, also providing new ideas and methods for the automatic evaluation of facial paralysis with small sample size.

Application of 3D Convolutional Neural Network

Ji and Xu et al. first proposed the 3D convolutional neural network.[21]. Now it has shown great advantages in the video, action recognition, facial expression[22][23][24] and other aspects.

Compared with 2D convolution, 3D convolution can better capture the time and space characteristics in the video. Because it operates on continuous frames, combines multiple continuous frames into a cube, and then uses the convolution kernel to calculate (as shown in Figure 9).

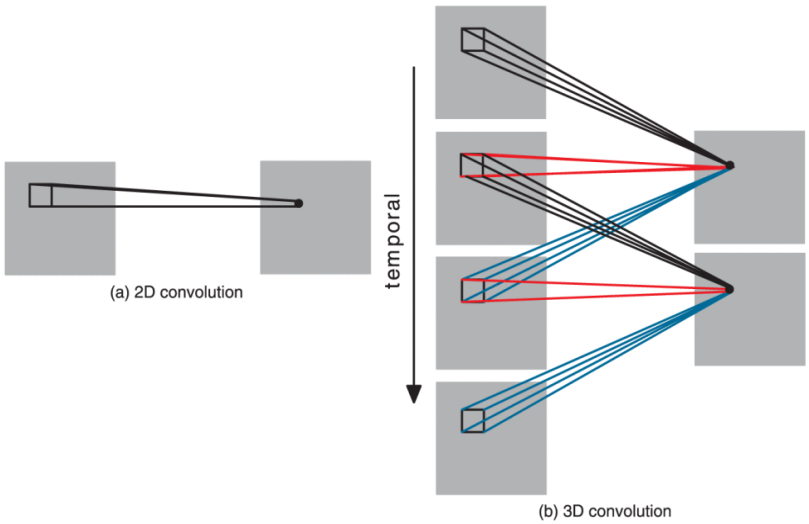


Fig.9. (a)2D convolution (b)3D convolution[21]

The 3D convolutional neural network performs 3D convolution kernel operations on the continuous frames of the video. This extracts the characteristics of the time and

space domain, which is more suitable for solving action recognition. This method is suitable for facial paralysis evaluation since the important criteria for facial paralysis evaluation is the motor ability and symmetry of facial muscles.

At present, 3D convolutional neural networks have been applied to the evaluation of facial paralysis. Storey and Jiang et al. [25] applied the 3D convolutional neural network to detecting and evaluating facial paralysis for the first time. They choose the H-B scale as the criterion and achieved an evaluation accuracy of 88%. This method weakens the problem of the insufficient sample size of facial paralysis to a certain extent through transfer learning and the method of introducing normal data. This shows that although the temporal and spatial characteristics are more suitable for facial paralysis assessment, the 3D accuracy rate still has the potential to improve. How to optimize the calculation speed and save time and cost still needs further research.

Considering the above research status at home and abroad, the current research on automatic evaluation of facial paralysis has the following problems:

① Most of them only focus on single, static or dynamic features that describe facial symmetry.

② Small sample size. Due to the protection of patient privacy and other reasons, a large number of authoritative public facial paralysis video data sets have not yet been constructed. Therefore, many current studies only assess the grade of facial paralysis for a single expression without considering the patient's overall facial condition, which cannot meet the needs of clinical diagnosis and evaluation.

The impact of the small sample size is relatively small for the traditional evaluation method by manually selecting features. The selection of feature points affects the calculation speed and accuracy of the classifier, but it does not have a good classification effect for adjacent levels of facial paralysis.

The neural network method has good accuracy on the problem of adjacent ratings of facial paralysis. However, it is still limited by the small sample size, and it is necessary to improve the effect of network training through methods such as sample enhancement.

③ Most of the current researches are based on static images as classified data. They rarely consider the impact of time-domain dynamic features on the overall evaluation. Since the facial paralysis data is in the form of video, this type of method may ignore the role of dynamic facial information in the study of facial paralysis

classification.

1.3 The Main Content and Innovations of the Topic

To overcome the shortcomings of current researches focusing on single-type feature methods, prevent the misjudgment of diagnostic information caused by single-feature methods and reduce the accuracy of the assessment. Using video data of facial paralysis patients, this thesis proposes a comprehensive facial paralysis detection and evaluation method based on facial landmark points and optical flow difference features.

The main work content is as follows:

① Preprocess the video data of patients with facial paralysis, perform sample enhancement, and extract keyframes: the extreme state of 7 types of expression images including no expression, staring, closed eyes, puffing, pouting, showing teeth, and frowning.

② Obtain the facial landmark points of the image, and extract the difference features of the left and right faces of the image accordingly; use the extracted features to train the SVM according to the H-B scale to realize the detection and classification of facial paralysis.

③ Extract optical flow features from 3 pairs of key facial regions-eyebrows, eyes, and mouth on facial expression images. Further, extract the optical flow differential features of the left and right facial regions based on facial symmetry, input them into the classifier and evaluate the facial paralysis data.

④ Integrate the evaluation results of the facial landmark method and the optical flow feature extraction method, and integrate the weighted dynamic and static features to improve the accuracy of facial paralysis evaluation.

The overall frame diagram is shown in Figure 10:

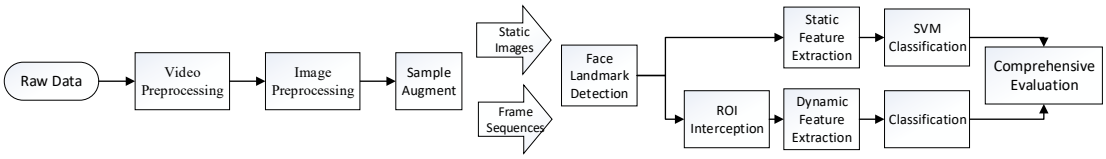


Fig. 10 The overall framework of facial paralysis comprehensive detection and evaluation based on facial landmark points and optical flow difference features

The main innovations of this thesis:

① Using the facial landmark method and the optical flow method, I extract the static and dynamic facial features simultaneously to improve the accuracy of classification.

② After image preprocessing, the keyframe of facial paralysis video data is automatically extracted.

1.4 Thesis structure

The writing of this thesis is divided into the following five chapters.

Chapter I: Introduction. This chapter explains the research background and significance of facial paralysis detection and evaluation, summarizes the current domestic and foreign research on the rating of facial paralysis, and analyzes the advantages and existing problems and deficiencies of the flash. Aiming at the problems existing in the research of facial paralysis automatic evaluation, this article's main research content and innovation points are described.

Chapter II: Video capture and preprocessing. First, the video acquisition conditions (environment, equipment requirements, pixels, frame rate, etc.), subject information (age, gender, and marked facial paralysis level), and the constructed data set are explained in Chapter II. The steps of data set preprocessing are described in detail, including rough video interception, video de-shake, video keyframe extraction, image normalization, which reduces the impact of the image on the classification effect and improves the processing speed. Finally, sample enhancement is adopted to solve the problem of a small sample size to a certain extent.

Chapter III: Static/dynamic feature extraction. This chapter introduces the facial landmark point determination algorithm in detail. The Ensemble of Regression Trees (ETR) algorithm is used to determine 68 facial landmark points. The static features of the keyframes are extracted considering the symmetry of the human face. As for extracting dynamic features, divide the facial expression image into ROIs based on the landmark points, extract the optical flow information of the three parts, and extract the optical flow difference features with the optical flow difference network.

Chapter IV: Classifier design, experimental plan and results. This chapter briefly describes the application of the SVM classifier and Long Short-Term Memory (Long Short-Term Memory) classifier in this thesis. The feasibility of the method proposed in this thesis is analyzed. Compared with traditional methods, it shows the innovative optimization points and shortcomings of the method proposed in this thesis.

Chapter V: Summarize the research content of this thesis, and look forward to the future research direction.

Chapter II Video Capture and Image Preprocessing

In order to realize the automatic assessment of facial paralysis, it is necessary to extract facial paralysis features that reflect facial symmetry. Before feature extraction, the video data needs to be preprocessed first. The purpose is to reduce the error, noise, redundancy and even errors in the original data so that the data is streamlined and reliable. It lays a good foundation for the subsequent feature extraction and classification work.

In this thesis, the facial paralysis video data undergo rough video interception, video stabilization, video keyframe extraction, and image normalization. All the preprocessing reduces the impact of the image itself on the classification performance. Finally, sample enhancement is used to solve the problem of a small sample size to a certain extent.

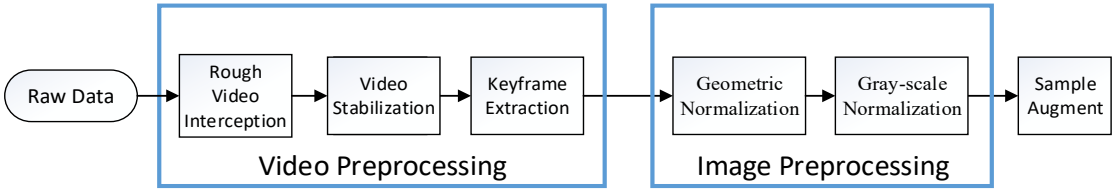


Fig. 11 Block diagram of video acquisition and image preprocessing flow

2.1 Video Capture

It is self-evident for research that collecting experimental data and establishing experimental data sets are important. A high-quality data set can often improve the quality of model training and the accuracy of prediction.

2.1.1 Video Capture Conditions

The doctor collects the videos in this thesis in various environments, including hospitals and subjects' homes. The lighting conditions in each collection environment are different. The video capture equipment is fixed, and the capture environment is indoors. The videos are divided into two groups. One video group has a size of 1280 pixels * 720 pixels, and the frame rate is about 25 frames per second, with a total of 10 cases. The other group has a size of 1920 pixels * 1080 pixels, and the frame rate is about 15 frames per second, with a total of 20 cases. The two groups of video images both have a 16:9 aspect ratio. The pixel density is the same, so the image size

is different.

2.1.2 Subject information

Since the videos used for facial palsy grading assessment involve the subjects' privacy, there is currently no public data set. In order to verify the effectiveness of the proposed method, our research team cooperated with the Rehabilitation Department of Huashan Hospital and the company to collect 25 videos of facial paralysis patients and 5 videos of normal subjects by ourselves.

Table 2 subject information

Subject No.	Gender	Evaluation date	H-B grade
1	M	2021/2/5	II
2	F	2020/12/16	III
3	M	2020/10/29	V
4	M	2021/4/13	V
5	F	2020/9/18	IV
6	F	2020/11/3	V
7	F	2021/2/19	V
8	F	2021/3/31	V
9	M	2021/3/12	IV
10	M	2020/7/29	V
11	M	2020/12/2	IV
12	F	2021/2/25	II
13	F	2020/7/24	III
14	M	2021/1/7	III
15	F	2021/2/7	V
16	F	2021/3/5	V
17	M	2020/9/14	II
18	F	2020/7/29	III
19	M	2021/3/18	V
20	F	2021/1/11	V
21	F	2019/12/4	III
22	F	2019/12/13	III
23	M	2019/11/28	IV
24	M	2019/12/5	IV
25	F	2019/12/12	IV
26	M	2021/4/10	I
27	M	2021/4/11	I
28	M	2021/4/10	I
29	M	2021/4/13	I
30	M	2021/4/13	I

In addition, the collected videos have been marked by professional medical staff. In gender classification, the data set contains 14 women and 16 men; By the level of facial paralysis, there are five cases of first-level, 3 cases of second-level, 6 cases of

third-level, 6 cases of fourth-level, and 10 cases of fifth-level.

The grading of facial paralysis images and videos are marked with the following 5 levels: I represents the first level of the H-B scale (normal people), choose to use II-IV to represent the second to the fourth level of the H-B scale rating, respectively. Because the six-level facial paralysis data is not collected in the data set, use V to indicate grade 5 or 6 facial paralysis on the H-B scale. The data labelling rule is that the label of a subject's video is the same.

2.2 Facial Paralysis Video Preprocessing

2.2.1 Rough Facial Video Interception

Videos are shoot in different environmental backgrounds, and thence there may be situations where background motion affects the keyframe capture, as shown in Figure 12. The thesis preliminarily crops the video image before capturing the keyframe. then, it takes the facial paralysis classification-related images as the interception target. Implementing this can eliminate the influence of irrelevant motion in the background on the feature extraction.

Fig. 12 Facial paralysis assessment video with background movement

The method selected in the thesis is to detect faces through the machine learning network dlib.get_frontal_face_detector. After face detection, I mark the center point of the face block diagram, as shown by the red dot in Figure 13; and set this point as the center to intercept the facial image. After the threshold test, I crop 150 pixels to the left and right and 200 pixels to the top and bottom. The size of the newly acquired video frame is 300*400 pixels.

After setting the center for the first time, let the threshold as 30 pixels. Mark the center of the face frame by frame. If the abscissa x or ordinate y of the current frame center deviates by over threshold from the calibration center, update the intercept center position. Otherwise, the center remains unchanged. As shown in Figure 13, the green dot is the center of the current frame, and the red dot is the center of the previous frame. The reason is that the center of face recognition is not stable, which may lead to incorrect keyframe recognition due to the continuous changes of the image interception during the keyframe capture.

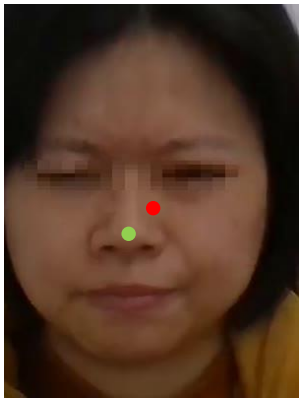


Fig. 13 Video image after the rough interception

Another advantage of rough face video capture is that the size of the video frame is reduced from 1280*720 or 1920*1080 to 300*400, which saves much computation for the subsequent video debounce and keyframe capture.

2.2.2 Facial Video Stabilization

In the process of video capture, there is random movement between consecutive image frame sequences, since the digital imaging equipment is affected by the carrier's posture change or the random vibration relative to the scene. The random movement will cause blurred images and unstable information, which greatly limits the effective utilization of image information[26].

Video image stabilization technology can solve this problem conveniently and efficiently. It has a wide range of applications in manual observation discrimination[27] and object detection, tracking[28] and compression[29].

A tripod, relatively stable, usually fixes the camera that records facial paralysis videos; however, the recording subject may unconsciously shake the head and body during the video recording process, so there is a problem of video shaking when capturing key facial information. For this reason, I apply 2D video stabilization technology to preprocess the video. It is based on extracting local features to estimate

the small, random-directed, high-frequency motion between frames and frames, then expand from local to global estimation, and achieve the goal of a stable and smooth image through compensation. There are generally four categories of methods: block matching method, projection method, feature method, and optical flow method[30]. The thesis chooses the optical flow method to stabilize the video image because the optical flow method is easy to implement. I intend to use the optical flow method to extract the dynamic features of the video in other parts of the thesis. The principle is to estimate the optical flow field based on the temporal-spatial image brightness gradient, which is carried out on the assumption that the gray level of the entire image is continuously changing.

The optical flow method has been applied in many fields, such as visual inertial tracking[31], moving target detection[32], motion intensity signal extraction[33], etc. The application of the optical flow method on motion detection in the field of view[34]is the focus of this thesis. The concept of optical flow is the projection of the motion of the object in the three-dimensional space on the two-dimensional plane. The instantaneous velocity of the moving pixel is projected by the spatially moving object on the observation imaging plane. It is produced by the relative speed of the object and the camera reflecting the moving direction and speed of the pixel corresponding to the object in a very small time[35]. The three elements of optical flow [35] are: motion velocity field; optical characteristic parts that carry motion information, such as color or gray-scale pixels; and imaging projections that make observations true, projecting from a three-dimensional scene to a two-dimensional plane.

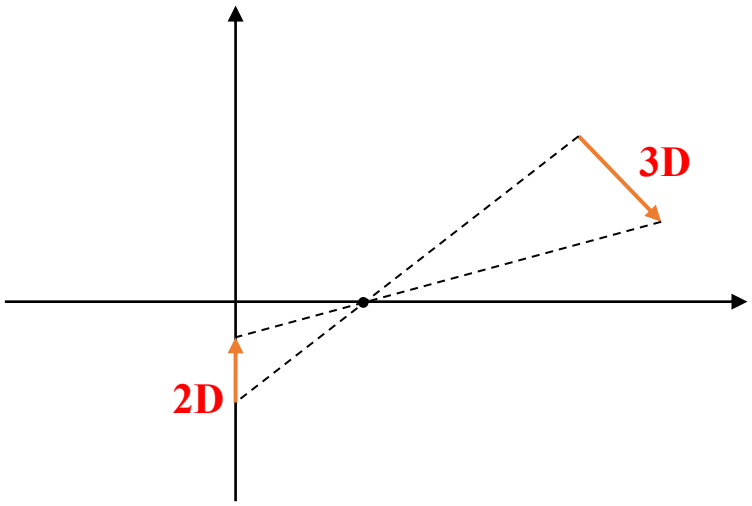


Fig.14 Projection of three-dimensional motion on a two-dimensional plane

Figure 14 shows that the three-dimensional space motion vector of the object is projected on the two-dimensional plane, and a two-dimensional vector describing the position change appear. When the movement interval is very small, it can be regarded as the instantaneous velocity vector of the point, the optical flow vector.

Two conditions need to be met for calculation using the optical flow method.

① The brightness is constant. If the brightness changes too much, it will misjudge the tracked two-dimensional pixels. Then, the same target between different frames cannot be found correctly.

② The changes in time will not cause drastic changes in the target position, and the displacement between adjacent frames is small.

For the basic optical flow calculation equation, consider this: Assuming $\boldsymbol{I(x,y,t)}$ is the light intensity of a pixel in the first frame. The pixel moves $(\boldsymbol{dx,dy})$ distance after a very short time $\boldsymbol{dt}$ in the second frame. Then $\boldsymbol{I(x+dx,y+dy,t+dt)}$ can express the light intensity of the pixel at this time. Because the light intensity of the pixel before and after the movement is unchanged, we can derive the formula (2.1):

$$I(x,y,t) = \mathrm{I(x + dx, y + dy, t + dt)} \quad (2.1)$$

Taylor expansion at the right end of equation 2.1, and get equation 2.2 after simplification:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}\frac{dt}{dt} = 0 \quad (2.2)$$

Suppose U and V are the velocity vectors of the optical flow along the x-axis and the y-axis:

$$u = \frac{dx}{dt}, v = \frac{dy}{dt} \quad (2.3)$$

The partial derivative of the light intensity of each pixel on the image along the x, y, and t directions is expressed as:

$$I_x = \frac{\partial I}{\partial x}, \ I_y = \frac{\partial I}{\partial y}, \ I_t = \frac{\partial I}{\partial t} \quad (2.4)$$

Final equation:

$$I_x u + I_y v + I_t = 0 \ (2.5)$$

There is only one constraint equation but two unknowns in the equation. In this case, the exact values of u and v cannot be obtained. Additional constraints need to be introduced. Different ideas lead to different constraints, resulting in different optical flow field calculation methods.

The optical flow method can be divided into the dense and the sparse optical flow by the density of the two-dimensional vector in the formed optical flow field.

In this thesis, the optical flow method de-jitter application is the Lucas-Kanade sparse optical flow method[37]. Because the continuous video frame contains much information, the LK optical flow method consumes fewer computing resources than the dense optical flow and relatively faster speed, which meets the requirements of video stabilization.

A set of points with common characteristics is the prerequisite for calculating LK sparse optical flow so that the results are relatively reliable and stable. Such as corner points or strong corner points[38]. The corner point was proposed by Harris[39]. It is where there are two maximum eigenvalues in the autocorrelation matrix of the second derivative of the image. This essentially means that there are at least two edges in different directions around this point. Strong corners refer to corners with strong robustness and are not easily interfered by noise.

To solve the problem of lacking constraint, the LK algorithm adds a spatially consistent hypothesis on the basis of the two assumptions of the original basic optical flow method: all adjacent elements have similar actions[40].

This shows that each pixel has the same optical flow vector in the M*M area around the target element. The LK optical flow method eliminates the ambiguity in the optical flow equation by considering the information of multiple adjacent pixels together. Compared with the point-by-point calculation method, the LK optical flow method is not sensitive to image noise.

Suppose that in a small neighborhood Ω range, the LK optical flow method estimates the optical flow vector by minimizing the weighted square sum of the neighborhood light intensity:

$$\sum_{(x,y\in\Omega)} \boldsymbol{W^2}(\boldsymbol{x})(\boldsymbol{I_x u} + \boldsymbol{I_y v} + \boldsymbol{I_t})^2 \quad \textbf{(2.6)}$$

$\boldsymbol{W^2}(\boldsymbol{x})$in formula 2.6 is the window weight function, which makes the center of the neighborhood weights more than the surrounding ones, reflecting that the optical flow information pays more attention to the central pixel.

For the points $x_1 \sim x_n$ in the neighborhood Ω, set:

$$\boldsymbol{V} = (u, v)^T, \boldsymbol{\nabla I}(X) = (I_x, I_y)^T \quad (2.7)$$

$$\text{Let } \mathbf{A} = (\boldsymbol{\nabla I}(X_1), \dots, \boldsymbol{\nabla I}(X_n))^T, \mathbf{A} = \begin{bmatrix} I_x(X_1) & I_y(X_1) \\ I_x(X_2) & I_y(X_2) \\ \vdots & \vdots \\ I_x(X_n) & I_y(X_n) \end{bmatrix} \quad (2.8)$$

$$\text{Let } \mathbf{W} = diag(\mathbf{W}(X_1), \dots, \mathbf{W}(X_n)), \mathbf{A} = \begin{bmatrix} \mathbf{W}(X_1) & 0 & 0 & 0 \\ 0 & \mathbf{W}(X_2) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{W}(X_n) \end{bmatrix} \quad (2.9)$$

$$\text{Let } \mathbf{b} = -\left(\frac{\partial I(X_1)}{\partial x}, \dots, \frac{\partial I(X_n)}{\partial x}\right)^T, \mathbf{b} = \begin{bmatrix} I_t(X_1) \\ I_t(X_2) \\ \vdots \\ I_t(X_n) \end{bmatrix} \quad (2.10)$$

Simplify by least squares method:

$$\mathbf{V} = -(\boldsymbol{A^T W^2 A})^{-1} \boldsymbol{A^T W^2 b} \quad (2.11)$$

The specific steps of the optical flow image stabilization algorithm in the thesis are:

① Read video frames sequentially and convert them to gray-scale images.

② Detect and track the corner points of the frame. First, use the function cv2.goodFeaturesToTrack to detect the corners of both previous and current frames. Adopt the function cv2.calcOpticalFlowPyrLK to calculate the optical flow of the small neighborhood Ω with the detected corners as the center. Apply the LK optical flow method above to track the corners of the gray-scale image. Find the rigid transformation $\boldsymbol{T}$ that maps the previous coordinate system with two sets of corner points to the current coordinate system. $\boldsymbol{T}$ contains only three parameters: dx, dy, da (angle)

③ Accumulate the estimated differential motion in step 2 to find the motion trajectory $\boldsymbol{f}$.

④ Smooth the trajectory $\boldsymbol{f}$ with an average window into the smoothed trajectory $\boldsymbol{f_{smooth}}$

⑤ Create a new transformation matrix, $\boldsymbol{T_{new}}$, from the previous frame to the current frame by finding the difference between the smooth trajectory and the original trajectory, and adding these differences back to the original transformation.

$$\boldsymbol{T_{new}} = \boldsymbol{T} + \boldsymbol{f_{smooth}} - \boldsymbol{f}$$

⑥ Apply the transformation matrix to the video.

Video stabilization can increase the accuracy and stability of the video frames. When the subject makes facial expressions as required, the video keyframe extraction will be misjudged due to the shaking of the body or equipment. After stabilization, the extraction of keyframes is more accurate. This lowers the proportion of redundant frames, thereby increasing the speed of subsequent feature extraction. For example, the number of keyframes from a video without video stabilization is 63. After video stabilization, the number of keyframes from the same video is 61 frames. The images

shown in Figure 15 are removed. These facial images are repeated expressionless images. The number of face images that need to be preprocessed can be reduced after reducing redundant keyframe extraction.

Fig.15 Redundant expressionless frames

2.2.3 Facial video keyframe capture

In video feature research, it is allowed to extract features from all frames of the facial paralysis video so that the video spatial-temporal features can be extracted relatively thoroughly. Its disadvantages are feature redundancy and a large amount of calculation, which leads to a decrease in algorithm speed. In the case of high requirements for the robustness of the extracted features, this method is a possible choice. But the application scenario of this work requires extracting facial paralysis video features in a relatively short time. So, I use the method of video keyframe feature extraction. Many full-frame sequences of facial paralysis videos do not contain related information. Keyframe extraction can greatly compress redundant video information, thereby increasing the rate and accuracy of video feature extraction by orders of magnitude.

Keyframes, also known as representative frames, are used to describe the key images of video shooting, which reflect the main content of video shooting. Keyframe is often used as the index of the video stream [41]. The extraction of keyframes, on the one hand, must reflect the main events of video shooting as completely and accurately as possible. On the other hand, the number of keyframes cannot be too large. The extraction algorithm should have lower computational complexity considering the amount of video data [41].

According to the criteria for facial paralysis grading, the facial expression at the maximum amplitude can better reflect the severity of facial paralysis. Therefore, it is necessary to extract the keyframes according to the strength of the expression action.

The greater the motion amplitude in the video, the more keyframes extracted; the smaller the motion amplitude, the fewer keyframes extracted. If the action is less than a certain threshold, it is regarded as static, and no keyframe extraction work is performed. The keyframes extracted in this way can reflect the degree of facial paralysis completely and accurately as well as the number is more appropriate.

This thesis extracts keyframes with the motion analysis method. The realization process is to: divide the video that has been roughly intercepted and stabilized into sub-videos with a length of 2 seconds. Analyze the optical flow of the object movement in each sub-video, and select the video frame with the lowest optical flow movements in the video as keyframe target; the keyframe is relatively static, corresponding to the largest or no expression.

The formula for calculating the motion amount of a video frame utilizing the optical flow method is as follows:

$$\boldsymbol{M(k)} = \sum\sum|\boldsymbol{L_x(i,j,k)}| + \left|\boldsymbol{L_y(i,j,k)}\right| \quad (2.12)$$

$\boldsymbol{M(k)}$ represents the amount of motion in the kth frame, $\boldsymbol{L_x(i,j,k)}$, $\boldsymbol{L_y(i,j,k)}$ respectively represent the x component and y component of the optical flow at the pixel point $\boldsymbol{(i,j)}$ of the kth frame. Take the local minimum value as the keyframe to be extracted.

$$\boldsymbol{M(k_i)} = \min[\boldsymbol{M(k)}] \quad (2.13)$$

Extracting the keyframes of facial paralysis video data can greatly reduce the amount of calculation in the subsequent process of feature extraction and data classification based on retaining the key information. Features will be closer to the standards of the H-B scale, which will greatly improve the performance of the facial palsy grading.

With the facial paralysis video undergoing the keyframe extraction, the keyframe sequence is preprocessed to determine the facial landmark points and divide the regions.

Figure 16-18 is the keyframe extraction effect. It shows that the keyframe has the largest expression amplitude, and the expression amplitude of the fifth frame before and after the keyframe is not as large as the selected keyframe.

Fig.16 (a) 5 frames before the stare keyframe

(b) stare keyframe (c) 5 frames after the stare keyframe

Fig.17 (a) 5 frames before the smile keyframe

(b) the smile keyframe (c) 5 frames after the smile keyframe

Fig.18 (a) 5 frames before frowning keyframe

(b) frowning keyframe (c) 5 frames after frowning keyframe

2.3 Image preprocessing

Image preprocessing plays a very important role in digital image processing. The

quality of image directly affects the image segmentation, classification, and recognition. The subsequent work on the location of facial key points and extraction of optical flow information in this study requires standard normalized images. However, the keyframe image has the problem of insufficient contrast. There may be offset and rotation of the position of the face, as well as differences in image ratio, color and brightness.

I carry out two preprocessing of geometric normalization and gray normalization on the face image to solve the above problems. Face correction and face cropping are the two steps of geometric normalization, for detecting landmark more accurately and eliminating useless figure information. The gray-scale normalization is to increase the contrast of the image and compensate illumination.

2.3.1 Geometric Normalization

The purpose of geometric normalization is to transform expression sub-images into uniform size, which is conducive to extracting facial features. It includes angle rotation and precise face interception.

In the process of facial paralysis video recording, even the frontal recording may have a slight tilt of the image, which will affect the performance of subsequent training and classification. In this thesis, the image is rotated and corrected according to the inclination angle of the line between the two eyes of the person to maintain the consistency of the face direction. This method does not require the secondary positioning of feature points. The specific correction principle is to determine the position coordinates of the pupil in the face image as e1 (x1, y1) and e2 (x2.y2), then the angle rotation formula of the image is shown in (2.14):

$$\boldsymbol{\theta} = \arctan\frac{y_2 - y_1}{x_2 - x_1} \quad (2.14)$$

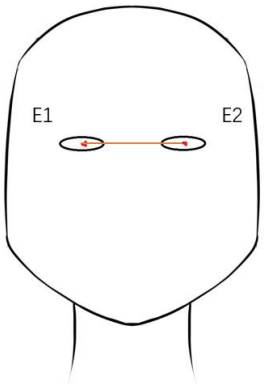


Fig.19 The pupil position and connection of the face image

According to the value of θ, the face image is rotated with the image center as the axis. When θ is less than 0, the image rotates counterclockwise, and when θ is greater than zero, it rotates clockwise. Finally, the pupil line E1E2 parallels with the horizontal line.

Fig.20 Before rotating (left) After rotating (right)

The data set is collected under the guidance of doctors. Faces are all facing the camera directly. The main error is the horizontal deflection of the face. The impact of head down/up and left-right tilt is very small and can be ignored.

On the basis of rough interception, I crop the rotated image more accurately, excluding areas such as the hair, neck and other areas that are not related to the extraction of facial features. Only keep facial areas with a consistent location. The specific cutting rules are as follows: assume the distance between the pupil centers of the two eyes to be d, the vertical upward cutting edge is 0.8d from the pupil, and the vertical downward cutting edge is 1.8d from pupil; left cutting border is horizontally 0.6d to the left of the left eye, and right cutting border is horizontally 0.6d to the right of the right eye. As shown in Figure 21

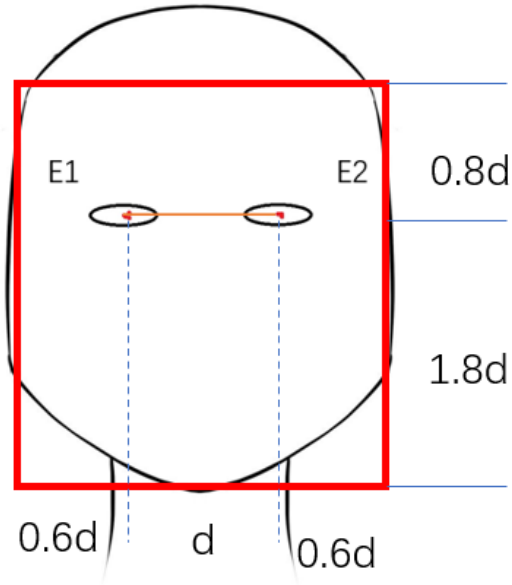


Fig.21 Cut rule

After the face is cropped, the size of all images is normalized to 330*390 pixels.

Fig.22 Fine-cropped face image

2.3.2 Gray Normalization

Because the videos of subjects in this thesis are collected in different environments, at different times, and under different lighting conditions, leading to a very unconcentrated facial image gray-scale distribution. There are big differences in the gray-scale distribution of different videos, directly affecting the subsequent feature extraction and classification. Therefore, it is necessary to normalize the gray level of the facial image to make the image details clearer and reduce the influence of the illumination angle and intensity. Common gray-scale normalization methods include mean variance normalization, gray-level change normalization, histogram

equalization, and so on.

I adopt the histogram mean method. The equalization of the histogram is a gray-scale transformation process: the current gray-scale distribution is transformed into an image with a wider range and a more uniform gray-scale distribution through a transformation function. The histogram of the original image is modified to be approximately uniformly distributed in the entire gray-scale interval. Therefore, the dynamic range of the image is expanded, and the contrast of the image is enhanced. The transformation function selected for equalization is usually the cumulative probability of gray-scale. The steps of the histogram equalization algorithm are as follows:

① Calculate the gray histogram of the original image $\boldsymbol{P}(\boldsymbol{S_k}) = \frac{\mathrm{n}_k}{n}$ (3.4), where n is the total number of pixels, and n_k is the number of pixels in the gray level $\boldsymbol{S_k}$

② Calculate the cumulative histogram of the original image $\boldsymbol{CDF}(\boldsymbol{S_k}) = \sum_{i=0}^{k} \frac{\mathrm{n}_i}{n} = \sum_{i=0}^{k} \boldsymbol{P}(\boldsymbol{S_i})$ (3.5)

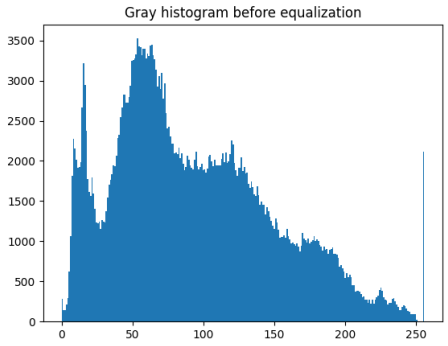
③ $\boldsymbol{D_j} = \boldsymbol{L} \cdot \boldsymbol{CDF}(\boldsymbol{S_i})$ (3.6), where $\boldsymbol{D_j}$ is the pixel of the target image, $\boldsymbol{CDF}(\boldsymbol{S_i})$ is the cumulative distribution of the source image gray level i, and L is the maximum gray level in the image (the gray level image is 255)

Fig.23

Fig.24

Fig.23 The image before the equalization of the gray histogram

Fig.24 The image after the equalization of the gray histogram



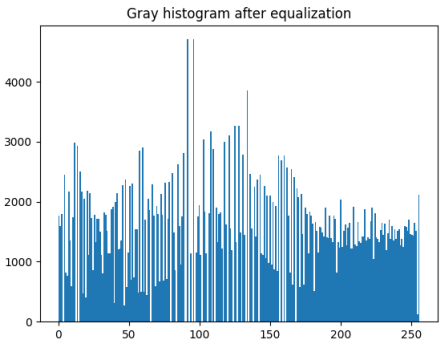


Fig.25 Gray-scale histogram before equalization Fig.26 Gray-scale histogram after equalization

After equalizing the gray-scale histogram, the problem is improved. The lighting environments of the left and right faces of the facial image differ, with the brightness of the left and right faces relatively more average. This makes the detection of facial landmarks more accurate, as shown in figure 28. The right eyebrow 23, 24 o'clock is accurately closer to the eyebrow.

Fig.27 Fig.28

Fig.27 Confirmation of the mark points at the front of the equalization

Fig.28 Confirmation of the mark points at the rear of the equalization

2.4 Data Set Construction

In this data set, I extract 1419 keyframes from 30 videos. The specific quantities of keyframe pictures for each facial paralysis level are shown in Table 3.

Table 3 Quantities of keyframe pictures for each level of facial paralysis

H-B grading	Number of samples	Number of pictures
I	5	247
II	3	143
III	6	296
IV	6	285
V	10	448

This thesis studies the grading effects of two different data set divisions:

① In-subject division: 80% and 20% of the extracted keyframe images are used as the training set and the test set, respectively. The keyframes collected by the same subject at the same time Frame images may be randomly divided into the training set and data set.

② Division among subjects: 1 video sample is used as the test set, and the other samples are used as the training set. The data collected by the same subject at the same time will all appear in the training set or test set.

2.5 Chapter Summary

This chapter details the steps of video preprocessing: video preprocessing and image preprocessing. Video preprocessing includes rough video interception based on face detection, video stabilization by LK optical flow method, and video keyframe extraction. Image preprocessing includes fine interception based on rough interception, geometric normalization and gray normalization, and finally sample enhancement.

Chapter III Static and Dynamic Feature Extraction

It is necessary to extract facial paralysis features reflecting the symmetry of the face aiming at realizing the automatic assessment of facial paralysis. The facial paralysis video is preprocessed, and the algorithm reduces the impact of the image itself on the classification effect. Then I extract the static and dynamic features of the image. This chapter details the extraction of the static and dynamic features of the image on the basis of determining the facial landmark points. First, determine 68 facial landmark points with the Ensemble of Regression Trees (ETR) algorithm. Then, extract the static features of the keyframes related to the symmetry of the human face. Finally, divide the facial expression area of the image on the basis of the landmark points, extract the optical flow information of three pairs of ROI, and draw the dynamic characteristics of the optical flow difference by the optical flow difference network.

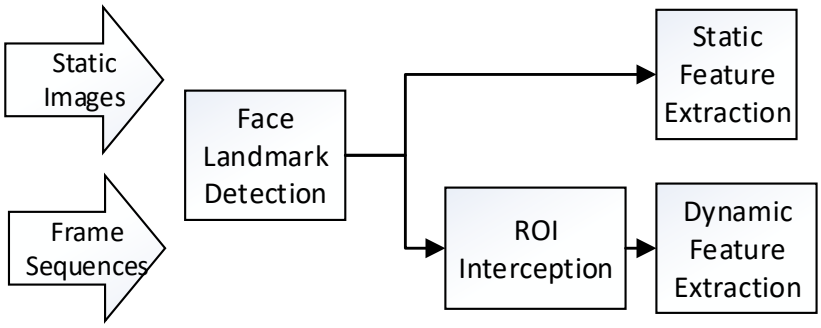


Fig.29 Block diagram of static/dynamic feature extraction

3.1 Facial landmark detection

Facial landmark detection is also called face positioning or alignment, which is to locate key areas of a given face image, such as eyebrows, eyes, nose, mouth, facial contours, etc. Facial landmark positioning occupies a very important position in computer vision. It is a basic problem in the field of computer vision and graphics. It is widely used in research on face recognition and verification, face exchange[42][43][44], etc. Its accuracy is directly related to the reliability of subsequent applications.

This thesis uses ETR to detect facial information and locate 68 key facial points. ETR algorithm uses cascaded regression trees to achieve face alignment. The core of the algorithm is the use of two-layer regression to build a mathematical model[45].

First, there is a training image data set $(\mathbf{I_1}, \mathbf{S_1}), \ldots, (\mathbf{I_n}, \mathbf{S_n})$, where $\mathbf{I_i}$ represents the i-th image in the set, $\mathbf{S_i}$ is the position of the landmarks on the face image.

In the first layer of regression training, the data form can be written as $(\boldsymbol{I_i}, \hat{\mathbf{S}}_{\boldsymbol{i}}^{(\mathrm{t})}, \Delta \mathbf{S_i}^{(\boldsymbol{t})})$, where $\boldsymbol{I_i}$ is the image in the data set, $\hat{\mathbf{S}}_{\boldsymbol{i}}^{(\mathrm{t})}$ is the predicted landmark position of the t-th layer in the first-level cascade regression, and $\Delta \mathbf{S_i}^{(\boldsymbol{t})}$ is the difference between the regression result of this layer and the true value.

The iterative process:

$$\hat{\mathbf{S}}^{(\mathrm{t}+1)} = \hat{\mathbf{S}}^{(\mathrm{t})} + \gamma_t(I, \hat{\mathbf{S}}^{(\mathrm{t})}) \quad (3.1)$$

$$\Delta \boldsymbol{S_i}^{(\boldsymbol{t})} = \boldsymbol{S_i} - \hat{\boldsymbol{S}}^{(t+1)} \quad (3.2)$$

Suppose the algorithm iterates continuously in this way. $\gamma_1 \ldots \gamma_k$ regressors will be produced when the number of first-level regression cascade layers is set to K layers. These K regressors are the regression models I need to obtain through training.

The input of the regressor γ_t is the current shape vector and training picture, and the output is the position update of all the landmarks. It can be obvious that in the first-level cascade regressor, every time the face landmarks pass a layer of cascade regressor, its positions will be updated once to reach a more accurate position.

The second level of regression, also a regression process inside γ_t, adopts the Gradient Tree Boosting Algorithm to generate a series of regression trees and finally completes the second level of regression. The object of the second level of regression is the difference between the true value and the current predicted value.

This thesis uses the ETR algorithm implemented in the dlib library to locate the landmarks of the face image. Use get_frontal_face_detector to determine the position of the face image in the image and shape_predictor to predict the position of each face key point.

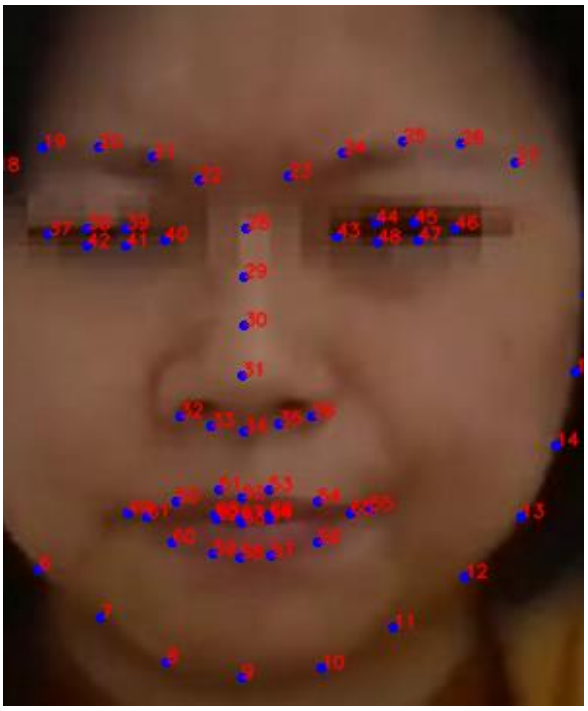


Fig.30 Location of facial landmarks

3.1.1 Facial region division

Taking into account the accuracy and pertinence of facial paralysis ratings and the complexity of optical flow feature extraction, I divide the faces of subjects into ROIs. Face image is divided on the basis of detected face landmarks into 3 groups of symmetrical left and right sides with the center line as the axis of symmetry: left and right eyebrows, left and right eyes, and left and right mouth. The expressions of subjects in all 6 areas are investigated.

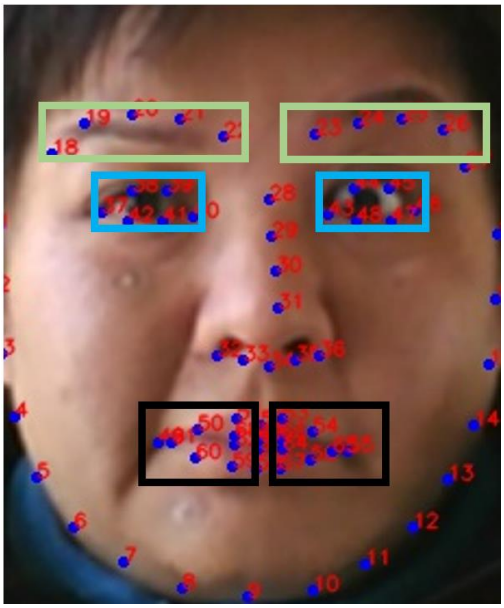


Fig.31 Division of the facial area

3.2 Feature Extraction Based on Facial Landmarks

The quantification of facial symmetry, extracting features based on facial landmarks, is based on 68 facial landmarks determined by the ETR algorithm. Comprehensively considering the H-B scale and video actions: a total of 32 features are selected to represent the severity of facial paralysis, such as the angle difference, circumference difference, and position difference of the corresponding vectors on the left and right sides of the eyebrows, eyes, and mouth, as shown in Figure 14.

All selected features below are normalized. p_i represents the i-th landmark point, y_i represents the ordinate of the i-th landmark point, $\angle \vec{a}, \vec{b}$ represents the angle between the two vectors, $\vec{h}$ represents the horizontal vector, $\overrightarrow{p_i p_j}$ represents the vector formed with p_i as the starting point and p_j as the ending point, and $\boldsymbol{L}(p_a \dots p_e)$ represents the length of the connection from p_a to p_e in order.

Eyebrow:

The angle between the eyebrow landmark vector and the horizontal vector:

$$\left||\angle \overrightarrow{p_{21}p_{20}}, \vec{h}| - |\angle \overrightarrow{p_{24}p_{25}}, \vec{h}|\right|$$

$$\left||\angle \overrightarrow{p_{22}p_{21}}, \vec{h}| - |\angle \overrightarrow{p_{23}p_{24}}, \vec{h}|\right|$$

$$\left||\angle \overrightarrow{p_{18}p_{19}}, \vec{h}| - |\angle \overrightarrow{p_{26}p_{27}}, \vec{h}|\right|$$

The corresponding height of eyebrow marking points 20, 19 and 25, 26:

$$\frac{|y_{20} - y_{25}|}{\boldsymbol{max}(y_{20}, y_{25})}$$

$$\frac{|y_{19} - y_{26}|}{\boldsymbol{max}(y_{19}, y_{26})}$$

Length of eyebrows 19~22 and 23~26:

$$\frac{|\boldsymbol{L}(p_{19}p_{20}p_{21}p_{22}) - \boldsymbol{L}(p_{23}p_{24}p_{25}p_{26})|}{\boldsymbol{max}(\boldsymbol{L}(p_{19}p_{20}p_{21}p_{22}), \boldsymbol{L}(p_{23}p_{24}p_{25}p_{26}))}$$

Eyes (Left eye: point 37~42; Right eye: point 43~48; 6 points each):

The difference between left and right eye circumference:

$$\frac{|\boldsymbol{L}(p_{37}p_{38}p_{39}p_{40}p_{41}p_{42}p_{37}) - \boldsymbol{L}(p_{43}p_{44}p_{45}p_{46}p_{47}p_{48}p_{43})|}{\boldsymbol{max}((p_{37}p_{38}p_{39}p_{40}p_{41}p_{42}p_{37}), \boldsymbol{L}(p_{43}p_{44}p_{45}p_{46}p_{47}p_{48}p_{43}))}$$

The height changes corresponding to point 38, 39 and 44, 45, the angle difference between the vector and the horizontal direction:

$$\frac{|y_{38} - y_{45}|}{\boldsymbol{max}(y_{38}, y_{45})}$$

$$\frac{|y_{39} - y_{44}|}{\boldsymbol{max}(y_{39}, y_{44})}$$

$$\left|\left|\angle\,\overrightarrow{p_{37}p_{38}},\vec{h}\right|-\left|\angle\,\overrightarrow{p_{45}p_{46}},\vec{h}\right|\right|$$

$$\left|\left|\angle\,\overrightarrow{p_{39}p_{40}},\vec{h}\right|-\left|\angle\,\overrightarrow{p_{43}p_{44}},\vec{h}\right|\right|$$

The height changes corresponding to point 41, 42 and 48, 47, and the angle difference between vector and horizontal direction::

$$\frac{|y_{42}-y_{47}|}{\boldsymbol{max}(y_{42},y_{47})}$$

$$\frac{|y_{41}-y_{48}|}{\boldsymbol{max}(y_{41},y_{48})}$$

$$\left|\left|\angle\,\overrightarrow{p_{37}p_{42}},\vec{h}\right|-\left|\angle\,\overrightarrow{p_{47}p_{46}},\vec{h}\right|\right|$$

$$\left|\left|\angle\,\overrightarrow{p_{41}p_{40}},\vec{h}\right|-\left|\angle\,\overrightarrow{p_{48}p_{43}},\vec{h}\right|\right|$$

Mouth：

The angle between the connecting vector of the mouth corner point 49 and 55 and the horizontal vector:

$$\left|\angle\,\overrightarrow{p_{49}p_{55}},\vec{h}\right|$$

The angle between the corner of the mouth and the center of the mouth point 49, 63 and 63, 65:

$$|\angle\,\overrightarrow{p_{49}p_{63}},\overrightarrow{p_{63}p_{65}}|$$

Symmetrical height difference between upper and lower lips:

$$\frac{|y_{50}-y_{54}|}{\boldsymbol{max}(y_{50},y_{54})}$$

$$\frac{|y_{51}-y_{53}|}{\boldsymbol{max}(y_{51},y_{53})}$$

$$\frac{|y_{60}-y_{56}|}{\boldsymbol{max}(y_{60},y_{56})}$$

$$\frac{|y_{57}-y_{59}|}{\boldsymbol{max}(y_{57},y_{59})}$$

$$\frac{|y_{61}-y_{65}|}{\boldsymbol{max}(y_{61},y_{65})}$$

$$\frac{|y_{62}-y_{64}|}{\boldsymbol{max}(y_{62},y_{64})}$$

$$\frac{|y_{68}-y_{66}|}{\boldsymbol{max}(y_{68},y_{66})}$$

Symmetrical circumference difference of upper and lower lips:

$$\frac{|\boldsymbol{L}(p_{49}p_{50}p_{51}p_{52})-\boldsymbol{L}(p_{55}p_{54}p_{53}p_{52})|}{\boldsymbol{max}(\boldsymbol{L}(p_{49}p_{50}p_{51}p_{52}),\boldsymbol{L}(p_{55}p_{54}p_{53}p_{52}))}$$

$$\frac{|\boldsymbol{L}(p_{49}p_{60}p_{59}p_{58}) - \boldsymbol{L}(p_{55}p_{56}p_{57}p_{58})|}{\boldsymbol{max}(\boldsymbol{L}(p_{49}p_{60}p_{59}p_{58}), \boldsymbol{L}(p_{55}p_{56}p_{57}p_{58}))}$$
$$\frac{|\boldsymbol{L}(p_{61}p_{62}p_{63}) - \boldsymbol{L}(p_{65}p_{64}p_{63})|}{\boldsymbol{max}(\boldsymbol{L}(p_{61}p_{62}p_{63}), \boldsymbol{L}(p_{65}p_{64}p_{63}))}$$
$$\frac{|\boldsymbol{L}(p_{61}p_{68}p_{67}) - \boldsymbol{L}(p_{65}p_{66}p_{67})|}{\boldsymbol{max}(\boldsymbol{L}(p_{61}p_{68}p_{67}), \boldsymbol{L}(p_{65}p_{66}p_{67}))}$$

Left and right inner and outer corners of the eyes to the corners of the mouth:

$$\frac{|\boldsymbol{L}(p_{37}p_{49}) - \boldsymbol{L}(p_{46}p_{55})|}{\boldsymbol{max}(\boldsymbol{L}(p_{37}p_{49}), \boldsymbol{L}(p_{46}p_{55}))}$$
$$\frac{|\boldsymbol{L}(p_{40}p_{49}) - \boldsymbol{L}(p_{43}p_{55})|}{\boldsymbol{max}(\boldsymbol{L}(p_{40}p_{49}), \boldsymbol{L}(p_{43}p_{55}))}$$

Left and right eyebrows and eyebrows to the corners of the mouth:

$$\frac{|\boldsymbol{L}(p_{22}p_{49}) - \boldsymbol{L}(p_{23}p_{55})|}{\boldsymbol{max}(\boldsymbol{L}(p_{22}p_{49}), \boldsymbol{L}(p_{23}p_{55}))}$$
$$\frac{|\boldsymbol{L}(p_{18}p_{49}) - \boldsymbol{L}(p_{27}p_{55})|}{\boldsymbol{max}(\boldsymbol{L}(p_{18}p_{49}), \boldsymbol{L}(p_{27}p_{55}))}$$

3.3 Optical Flow Difference Feature Extraction based on Image Data

In the process of a doctor's diagnosis of a patient with facial paralysis, they ask the patient to make a variety of specific facial expressions. The diagnosis relies much on observing the facial muscles' movement ability when the expression is at its maximum.

Therefore, this thesis selects two types of images: the patient's state of expressionlessness and extreme expression. Then I calculate the optical flow characteristics of the two images using static optical flow calculations for the automatic rating of facial paralysis.

This thesis chooses Horn-Schunck dense optical flow method for optical flow feature extraction. Compared with sparse optical flow, the dense optical flow does not select some corner points in the image (usually corner points) for calculation[46]. It matches the image pixel by pixel and calculate the offset of all points. And then get the optical flow field so as to carry out the registration. Therefore, the calculation amount will be significantly greater than that of sparse optical flow. However, the effect is generally better than that of sparse optical flow. This thesis only extracts the optical flow characteristics of the three groups of ROIs of face, which can reduce the data amount calculated by the optical flow to an acceptable level. And the clinical

evaluation of facial paralysis needs to be considered as a whole, so the optical flow of the image point by point needs to be considered. This is closer to clinical diagnosis.

The HS optical flow algorithm introduces global smoothing constraints to estimate the motion in the image. The facial optical flow information is based on the basic optical flow constraints. The HS optical flow method adds additional conditions: set the speed of the pixels in the image and its neighboring elements to be similar or the same, and the speed change at each place in the optical flow field is smooth[47].

The smoothing constraint which can be expressed as the symbol definition in the formula is the same as that of the optical flow formula in Chapter II:

$$\nabla^2 u = \frac{\partial^2 u}{\partial^2 x} + \frac{\partial^2 u}{\partial^2 y}, \nabla^2 v = \frac{\partial^2 v}{\partial^2 x} + \frac{\partial^2 v}{\partial^2 y} \quad (3.3)$$

Approximate $\nabla^2 u$, $\nabla^2 v$:

$$\nabla^2 u \cong \kappa\left(\bar{u}_{i,j,k} - u_{i,j,k}\right), \nabla^2 v \cong \kappa\left(\bar{v}_{i,j,k} - v_{i,j,k}\right) \quad (3.4)$$

K represents the range of the neighborhood, and the average value of the neighborhood is expressed as [48].

$$\bar{u}_{i,j,k} = \frac{1}{6}\left\{u_{i-1,j,k} + u_{i,j+1,k} + u_{i+1,j,k} + u_{i,j-1,k}\right\} + \frac{1}{12}\left\{u_{i-1,j-1,k} + u_{i-1,j+1,k} + u_{i+1,j+1,k} + u_{i+1,j-1,k}\right\} \quad (3.5)$$

$$\bar{v}_{i,j,k} = \frac{1}{6}\left\{v_{i-1,j,k} + v_{i,j+1,k} + v_{i+1,j,k} + v_{i,j-1,k}\right\} + \frac{1}{12}\left\{v_{i-1,j-1,k} + v_{i-1,j+1,k} + v_{i+1,j+1,k} + v_{i+1,j-1,k}\right\} \quad (3.6)$$

We can establish the energy function with two constraints:

$$E(u,v) = \iint\left[(I_x u + I_y v + I_t)^2 + \alpha^2(\|\nabla u\|^2 + \|\nabla v\|^2)\right]dxdy \quad (3.7)$$

α is the smoothing weight coefficient. The larger the value is, the smoother the optical flow is. The goal is to calculate the value of ∇u and ∇v when the energy function E gets the minimum value. A reasonable optical flow estimation should be to make the optical flow field of the above two factors as small as possible and solve the functional extremum problem through the Lagrange equation to derive equations, where Δ is the Laplace operator:

$$I_x(I_x u + I_y v + I_t) - \alpha^2 \Delta u = 0$$

$$I_y(I_x u + I_y v + I_t) - \alpha^2 \Delta v = 0 \quad (3.8)$$

Finally, the system of equations can be solved by iterative method：

$$u^{k+1} = \bar{u}^k - \frac{I_x(I_x\bar{u}^k + I_y\bar{v}^k + I_t)}{\alpha^2 + {I_x}^2 + {I_y}^2}$$

$$v^{k+1} = \bar{v}^k - \frac{I_x(I_x\bar{u}^k + I_y\bar{v}^k + I_t)}{\alpha^2 + {I_x}^2 + {I_y}^2} \quad (3.9)$$

k represents the number of iterations. The initial values of optical flow $u(0)$ and $v(0)$ are generally 0. When the difference between before and after the iteration is less than the given threshold, the iteration ends[49].

3.3.1 Static Facial Optical Flow Features

Extract optical flow features of relevant facial regions for different actions, and calculate the optical flow of the face without expression state and the maximum state of each expression. I extract three groups of regions under different expressions at the same time and acquire ROI optical flow information between the expressionless state and the maximum state by the optical flow method.

Fig.32 Optical flow of the left eyebrow region

Fig.33 Optical flow of the right eyebrow region

3.3.2 Symmetrical Optical Flow Difference Features

For different facial actions, I compare and analyze the optical flow difference characteristics of the relevant areas on the left and right sides of the face. The optical flow image of the left face is subtracted from the optical flow image of the right face after the mirror transformation processing, whose result is the optical flow difference image. The optical flow difference formula is as follows:

$$\boldsymbol{W(D)} = \boldsymbol{L(W)} - \boldsymbol{R(W)} \quad (3.10)$$

$\boldsymbol{W(D)}$ represents the optical flow difference. $\boldsymbol{L(W)}$ represents the optical flow information after mirror transformation on the left side of the face. $\boldsymbol{R(W)}$ represents the optical flow information on the right side of the face. By comparing the optical flow information of the left face and the right face, the difference information of the optical flow is obtained. Fig.34 is the visualized image about the difference of the eyebrows optical flow.

Fig.34 Diagram of the difference of eyebrow light flow

3.4 Chapter Summary

This chapter details the static and dynamic feature extraction after video preprocessing. For static features, I use ETR to determine the facial feature points, and divide the face image according to the distribution of the feature points. I then introduce the selection of features based on the symmetry of the left and right parts of the face and list the calculation method of 32 features. For dynamic features, I use expressionless state and maximum state images as data. I then extract optical flow information from the optical flow difference network to extract optical flow difference characteristics.

Chapter IV Classifier Design, Experimental Scheme and Results

This chapter introduces facial paralysis classification based on static features and dynamic features of chapter three. The 32-dimensional static feature vector is input into the SVM classifier for training and testing; the dynamic optical flow differential image is first converted into a high-dimensional matrix, and then the dimensionality is reduced by Principal Component Analysis (PCA), and finally, a 60-dimensional feature vector is formed as the input of the LSTM network for train and test.

4.1 Experiment Settings

In this thesis, I mark the facial paralysis grades I to V as positive integers 1 to 5 as the data labels.

The static feature classifier of the experiment is SVM.

In-subject division: divide 80% of all data into the train set, 20% into the test set. All keyframes of the same subject in the same video are randomly distributed into train and test sets according to the ratio. The five-fold cross-validation method is used to optimize the SVM kernel function and parameters on the training set: divide the data into 5 randomly, and four of them are taken as the training set for training to obtain the classifier, and then use the learned classifier to test the remaining data. Loop 5 times and select classifiers, functions and parameters with high average accuracy and small variance under different training sets.

Division among subjects: set the data of 29 subjects as the train and the data of 1 subject as the test each time. Repeat the loop to obtain the predictive rating results of 30 subjects.

The dynamic feature classifier is LSTM. The division of training set and test set is the same as SVM.

4.2 SVM Classification based on Static Features

SVM is a machine learning method based on the VC dimension theory of statistical learning and the principle of structural risk minimization. It shows many unique advantages in solving small sample, nonlinear and high-dimensional pattern recognition problems, and to a large extent, overcomes the problems of "dimension

disaster" and "over-learning"[50]. Due to the limited objective conditions, this thesis only collects 30 subjects' videos and intercepts 1419 keyframe images. The number of facial paralysis images of certain grades is only 143, which is a typical small-sample study. Therefore, SVM is chosen as the classifier.

SVM is developed from the optimal classification surface under the linearly separable case, and the basic idea can be explained by two types of linearly separable cases**Virhe. Viitteen lähdettä ei löytynyt.**. As shown in Figure 35, the blue dots and orange dots represent two types of samples. A hyperplane can classify the two kinds of the samples to different sides. And this is the linear result of a linear core SVM classifier.

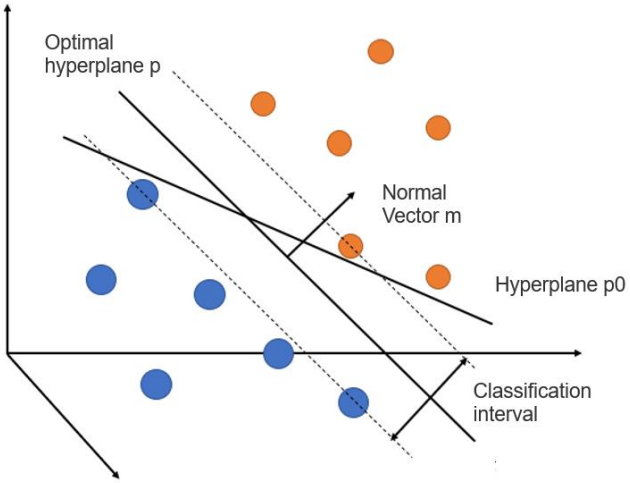


Fig.35 SVM hyperplane determination

There are countless such hyperplanes in space. The target to find the optimal solution is to find a hyperplane that can separate the two types of training samples and maximize the geometric distance between each sample and the hyperplane.

Suppose this n-dimensional hyperplane can be expressed by equation $\boldsymbol{\varpi}x + b = 0, \boldsymbol{\varpi}\epsilon\boldsymbol{R}^n, b\epsilon\boldsymbol{R}$.

Then in order to find this n-dimensional hyperplane, we can express the condition of the maximum geometric interval as: the geometric interval $\gamma_i = y_i(\frac{\boldsymbol{\varpi}}{\|\boldsymbol{\varpi}\|} \cdot x_i + \frac{\boldsymbol{b}}{\|\boldsymbol{\varpi}\|})$ of sample points(x_i, y_i) on the hyperplane is the largest.

Maximizing γ is equivalent to maximizing $\frac{1}{\|\boldsymbol{\varpi}\|}$ is equivalent to minimizing $\frac{\|\boldsymbol{\varpi}\|^2}{2}$, this constraint can be simplified to ：

$$y_i(\boldsymbol{\varpi} \cdot x_i + b) \geq 1, i = 1,2, \dots, N \quad (4.1)$$

In practical applications, there are many practical problems that cannot be used to classify the training set well through the linear hyperplane. In order to solve this

problem, the concept of the soft interval is introduced, and a slack variable ξ is introduced to indicate the degree of dissatisfaction of the sample to the constraint. Each sample has a corresponding slack variable.

We can transform the nonlinear sample classification problem into:

$$\min_{\boldsymbol{\varpi},\boldsymbol{b},\boldsymbol{\xi}_i} \frac{1}{2}\|\boldsymbol{\varpi}\|^2 + C\sum_{i=1}^{m}\boldsymbol{\xi}_i \quad (4.2)$$

$$s.t.\ y_i(\boldsymbol{\varpi}\cdot x_i + b) \geq 1 - \boldsymbol{\xi}_i \quad \boldsymbol{\xi}_i \geq 0, i = 1,2,\ldots,N \quad (4.3)$$

C is a penalty function. The larger the C, the greater the penalty for classification[51].

Nonlinear classification problem in the input space can be transformed into a linear classification problem in a certain dimensional feature space by nonlinear transformation, learning linear SVM in the high-dimensional feature space[50]. Both the objective function and the classification decision function only involve the inner product between the instance and the instance in the dual problem of linear SVM learning. So, there is no need to explicitly specify the nonlinear transformation. The kernel function can deal with the nonlinear work.

SVM is a commonly used method of facial paralysis feature classification, and utilizing SVM can also make a horizontal comparison with existing research. Therefore, I choose SVM as the static feature classifier.

4.2.1 Model Parameter Tuning

The kernel function is one of the few parameters that can be adjusted in SVM, including linear kernel function, polynomial kernel function, radial basis kernel function, and sigmoid kernel function. The choice of kernel function includes the choice of kernel function type and the choice of parameters after determining the kernel function type. This thesis uses five-fold cross-validation to select the kernel function and parameters. Divide the data into 5 groups at random and take four of them each time as the training set. The remaining 1 group as the test set. Choose the model with the best classification accuracy in the final test set. Table 4 shows the classification accuracy of the optimal model with different kernel functions selected by SVM.

Table 4 Comparison results of kernel functions

Accuracy/Kernel Function	Linear %	Polynomial%	Radial basis%	Sigmoid%
Accuracy	68.42	84.21	47.37	31.58

For the polynomial kernel function, the default value of d is 3, which means the highest degree of the polynomial kernel function, the most critical parameter. Table 5 shows the relationship between the highest order term of the polynomial kernel function and the accuracy. The default value of the gamma parameter in the kernel function is 1/k, where k represents the number of categories and is represented by g. I divide facial paralysis into 5 categories according to the severity of facial paralysis, then g=1/5. The default value of coef0 in the kernel function is 0, which is represented by r. The Gamma parameter and coef0 both remain default values without adjustments.

Table 5 Comparison result table of parameter d

parameter d	1	2	3	4	5	6	7	8
accuracy%	68.42	78.95	84.21	84.12	83.21	78.95	73.68	66.34

4.3 LSTM Classification based on Dynamic Optical Flow Features

The process of extracting dynamic optical flow features in this thesis is to obtain optical flow information through an optical flow algorithm, mirror the right partial facial area, and subtract the left facial partial and the right facial partial optical flow to obtain the optical flow difference image. Convert each optical flow difference image into a high-dimensional matrix, and then use PCA to reduce the dimension, and finally form a 60-dimensional feature vector input to the LSTM network for classification.

LSTM is very popular in processing time-dependent data. It has achieved remarkable results in sentiment analysis, machine translation, and text generation. The dynamic optical flow information is correlated highly with time. The adopted method is to classify the optical flow sequence, so I choose LSTM to achieve classification.

LSTM is a special recurrent neural network RNN, mainly to solve the problem of gradient disappearance and gradient explosion in the long training process. LSTM can solve complex and artificial long-lag tasks that have never been solved by previous recursive network algorithms[53].

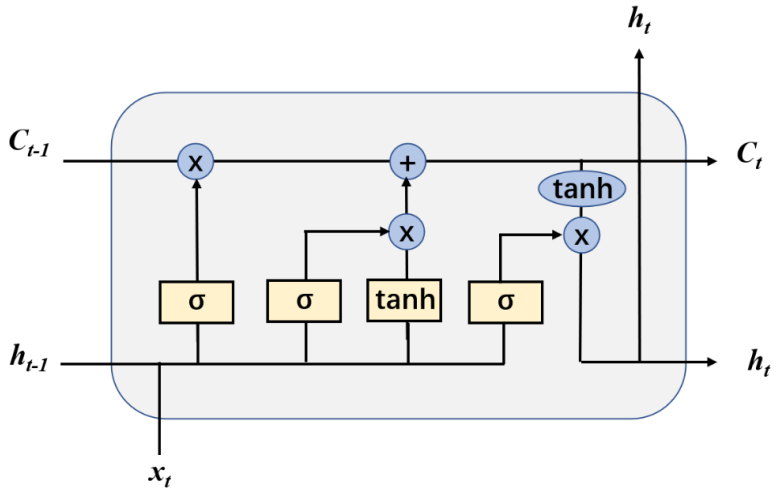


Fig.36 LSTM cell basic structure diagram

LSTM has a chain form of repeating neural network modules, and the basic structure of one neural network module is shown in figure 36. Each cell of LSTM has three gates: input gate, forget gate, and output gate. In actual application, I don't care about the state of the cell itself. Take $\boldsymbol{x_t}$ as input, and take its present state $\boldsymbol{h_t}$ as the final output.

Among the three groups of facial regions, there are local regions that contain local facial details. PCA calculation of the same image is independent and in parallel, and the corresponding feature vectors F1, F2, and F3 are extracted, respectively corresponding to the optical flow features of the three different areas of the face. After that, these features are multiplied by corresponding weighting coefficients (α1, α2, and α3), and the fused feature vector F is obtained. The 60-dimensional vector F is sent to the LSTM network, combined with the softmax layer to predict the class label. I get the result of the classification evaluation at last.

LSTM hyperparameter settings input_size = 60, hidden_size = 120 , num_layers = 2, batch_size = 1, learning_rate = 0.0004 and keep_prob = 1.0

4.4 Voting Ensemble Classifier

A voting ensemble classifier is a method of combining multiple models into one model. This thesis combines the two classifiers, SVM and LSTM, for voting.

There are usually two ways of voting integration:

① Hard voting. This means that each classifier gets a vote and then counts all votes to see what the winning class is. Weights can also be assigned to classifiers, which may give some classifiers more votes than others.

② Soft voting. Add the predicted probabilities of each classifier, and then select

the most probable class from them. This is a method recommended by the scikit-learn documentation and is suitable for well-calibrated classifiers. Each classifier can add weights.

This thesis chooses the soft voting method, which assigns different weights to the static feature classification and optical flow feature classification according to the facial paralysis prediction results. Since static feature classification is more effective for class IV and class V, the weight is higher when the static feature classification prediction is set to class IV and class V; dynamic feature classification is better for class II and class III classification, and the dynamic feature classification prediction is set as The weight is higher at II and III level. After adding the weights, the predicted probabilities are added together to get the most likely facial paralysis grading result.

$$\boldsymbol{p}_{\boldsymbol{final}} = \boldsymbol{w}_{\boldsymbol{stable}}\boldsymbol{p}_{\boldsymbol{stable}} + \boldsymbol{w}_{\boldsymbol{motive}}\boldsymbol{p}_{\boldsymbol{motive}} \quad (4.4)$$

4.5 Facial Palsy Grading Results

This thesis evaluates the classification effect by accuracy, precision, recall and F1-Measure indicators[54]. In the specified test data set, the accuracy is the ratio of the number of samples correctly classified by the classifier to the total number of samples in the data set; the precision refers to the ratio of the correct predictions being positive to all the predictions being positive. In other words, it is how many samples are correctly classified in the results of the predictions that are positive samples; If the sample is balanced, the values of accuracy and precision represent the classification effect. The larger the value, the better the effect. Recall rate refers to the proportion of correct positive predictions to all actual positives, which is the correct classification of the results of the true positive sample; F1 is the average of the precision rate and recall rate of a comprehensive index. The larger the F1 is, the better the classifier performance is. Table 6 shows the statistics of the number of misclassified and matched classes of the confusion matrix classification model. If the model is predictive, it is better to get more true positive (TP) samples and true negative (TN) samples. On the contrary, the less false positive (FP) and false negative (FN) samples are better.

Table 6 Confusion matrix

Forecast/actual	**Actual positive**	**Actual negative**
Positive case	True Positive	False Positive
negative case	False Negative	True Negative

The accuracy rate can be calculated by the formula (4.4):

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+FP+FN+TN}} \qquad (4.4)$$

From the concepts of precision rate and recall rate, we can derive formula (4.5) and formula (4.6):

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \qquad (4.5)$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \qquad (4.6)$$

The weighted average of Precision and Recall is also called F1 value, as shown in the following formula (4.7):

$$\boldsymbol{F1} = \frac{\boldsymbol{2Precision*Recall}}{\boldsymbol{Precision+Recall}} \qquad (4.7)$$

The evaluation of the experimental effect needs to be considered from both directions of the positive classification index and the negative classification index. Two methods of macro average and micro average are commonly used. The macro average treats each category equally, counts the index value of each category, calculates the accuracy, recall and F1 value of each category separately, and averages the index values. Micro-averaging starts from all categories, does not distinguish between categories, and performs statistics on each instance in the data set without classification to establish a global confusion matrix, calculates the corresponding proportion, and treats all samples equally.

Tables 7 and 8 show the results of the 5-fold cross-validation test for extracting keyframes.

Table 7 The accuracy rates of the static feature + SVM facial paralysis evaluation

H-B grading	**Completely match (%)**	**one grade difference (%)**	**two grades difference (%)**	**three grades difference (%)**
I (normal)	92.5	3.3	4.2	0
II	30.7	60.4	8.7	0.2
III	65.4	15.4	15.2	4
IV	50.9	30.1	19.5	4.2
V	75.7	23.1	1.2	0

Table 8 The accuracy rates of the optical flow + LSTM facial paralysis evaluation

H-B grading	Completely match (%)	one grade difference (%)	two grades difference (%)	three grades difference (%)
I (normal)	94.3	4.3	2.4	0
II	28.7	55.4	13.7	2.2
III	78.4	8.4	12.3	0.9
IV	40.9	50.1	8.5	0.5
V	70.5	28.5	1.2	0

The comprehensive accuracy rate of the two methods is below 70%, and the model does not reach a satisfactory grading effect.

Considering that the degree of facial paralysis should comprehensively consider all the expressions of the entire video, this thesis designs a second training method: 1 video is used as the test set, and the other videos are used as the training set. The facial paralysis rating is predicted by weighing judgments on all facial expressions belonging to the same patient in the same video data.

I choose Gabor+SVM[55], LBP+SVM[56], optical flow (image)+SVM[57] methods as comparative experiments.

I adopt the accuracy of the micro-average method, and the final accuracy results are as follows.

Table 9 Comparison of facial paralysis grading results by different methods

Methods	Accuracy	precision	recall	average F1
Gabor+SVM [55]	70%	68.97%	69.74%	68.82%
LBP+SVM[56]	62.7%	63.1%	63.1%	62.91%
optical flow+SVM[57]	84.21%	84.52%	85.42%	84.74%
Static features+SVM	93.33%	90%	90%	90%
Optical flow features +LSTM	86.67%	89.29%	86%	86.31%
Static features + Optical flow features + Voting Ensemble classification	93.33%	94.29%	91.33%	91.87%

The confusion matrix of the static feature + SVM method is shown in Figure 37:

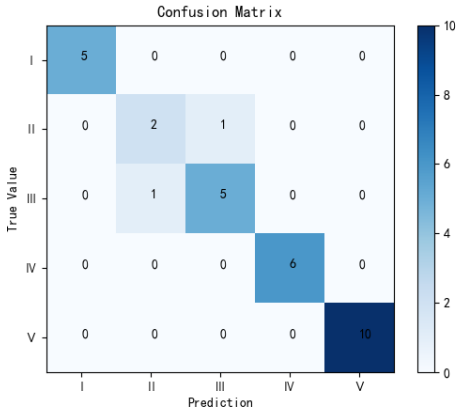


Fig.37 Confusion matrix of static feature + SVM method

The confusion matrix of the optical flow feature + LSTM method is shown in Figure 38:

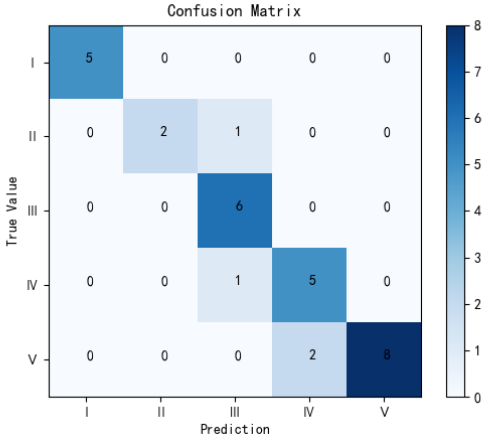


Fig.38 Optical flow feature + confusion matrix of LSTM method

The confusion matrix of static feature + optical flow feature + voting ensemble classification method is shown in Figure 39:

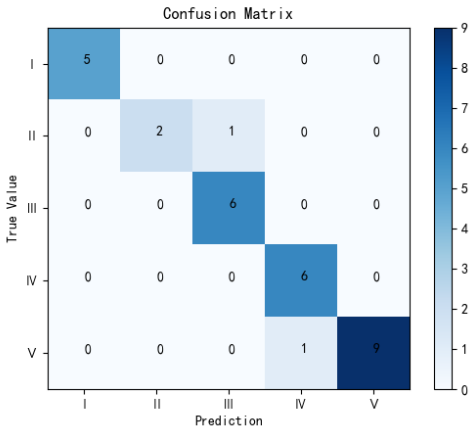


Fig.39 Static feature + optical flow feature + voting ensemble classification method

The accuracy of the three methods in this thesis has been greatly improved from 70% to 90% compared with the classification accuracy of[55][56][57]. The static feature + SVM method is more accurate for the classification of grade IV and V severe facial paralysis, and the optical flow feature + LSTM is more accurate for the classification of grade II and III severe facial paralysis, but the overall accuracy rate is 6.67% lower than that of the static feature + SVM method.

Compared with the optical flow feature + LSTM method, the static feature + optical flow feature + voting ensemble classification method has an increase of 6.67% in accuracy. The accuracy rate, recall rate, and f1 are all improved by about 5%, which reduces the error of grade IV and V facial paralysis. Judgment probability.

The final accuracy of the static feature + optical flow feature + voting integrated classification method is the same as that of the static feature extraction + SVM method alone, but the accuracy rate is increased by 4.29%, the recall rate is increased by 1.33%, and f1 is increased by 1.87%. It reduces the probability of misjudgment of grade III facial paralysis but increases the probability of misjudgment of grade V facial paralysis.

The performance of the static feature + optical flow feature + voting ensemble classification method has been improved. The reason is that for the data predicted to be level II and III facial paralysis, the weight of their static feature is reduced, and the weight of their optical flow feature is increased in the weighted voting. For data predicted to be grade IV and grade V facial paralysis, the weight of static features increases, and the weight of optical flow features decreases. In this way, I combine the advantages of the static feature + SVM method and the optical flow feature + LSTM method, increase the accuracy of the third-level facial paralysis with a small sample proportion, and reduce the accuracy of the fifth-level facial paralysis with a large sample proportion, resulting in the improvement on precision rate, recall rate and f1.

The method in this thesis combines static features and optical flow features to achieve a comprehensive grading assessment of facial paralysis. It further improve the performance of automatic grading and assessment of facial paralysis. Through adjusting the parameters, you can see that the accuracy of the model, the recall rate, and f1 are all improved by about 3%, and the comprehensive evaluation accuracy rate reaches about 93%.

4.6 Chapter Summary

This chapter introduces the choice of experimental classifiers. Static feature

classifier chooses to use SVM, SVM parameters and kernel functions are tuned through the five-fold cross-validation method, and dynamic feature classifier chooses LSTM. Finally, the two classifiers, SVM and LSTM, vote to obtain the final facial paralysis grade prediction. Compared with the existing methods, the performance of the method in this thesis has been significantly improved. Among all three methods in this thesis, the static feature + optical flow feature + voting integrated classification method combines the advantages of dynamic and static features classification. Its accuracy, precision, recall rate and f1 are the highest among the three methods.

Chapter V Summary and Prospect

5.1 Summary

The facial paralysis grading assessment is of great significance to the optimization of the treatment plan and prognosis of patients with facial paralysis. The traditional method is based on the subjective analysis of clinicians to grade facial paralysis. The evaluation results obtained are affected to a certain extent by the experience of clinicians. The evaluation results of facial paralysis grade are not accurate, which affects the formulation of treatment plans for patients with facial paralysis.

Most facial paralysis recognition methods based on computer vision technology only uses overall information to evaluate facial paralysis, which is not targeted and not comprehensive enough. This thesis proposes a facial palsy grading assessment method based on facial landmark feature extraction and optical flow difference feature. This method combines static facial landmark feature and optical flow information of dynamic video data to evaluate facial palsy.

The main work is as follows:

① Facial paralysis video preprocessing. A facial video sequence is drawn that can be identified by landmarks.

② Keyframe image extraction. Extract images of the neutral state of the face and the state of maximum facial motion of the subject with different expressions. And perform preprocessing and region division on the obtained images.

③ Facial paralysis static feature extraction. Extract the facial paralysis features representing the left and right symmetry of the face depending on the determined key points of the face. Input the features into the SVM classifier to obtain the facial paralysis rating results based on static features

④ Facial paralysis dynamic feature extraction. Obtain the optical flow through the images of the state of no expression and the state of maximum expression, calculate the optical flow difference features of the corresponding areas on the left and right sides of the face, and predict the facial paralysis grading based on dynamic features.

⑤ Comprehensive evaluation of facial paralysis classification. Evaluation results are based on artificial feature extraction and optical flow difference features.

Experiments show that the method proposed in this thesis improves the accuracy of facial paralysis grading assessment compared with previous methods. It also provides assistance for doctors to make reasonable judgments and treatments for patients.

5.2 Prospect

In this thesis, I adopt the optical flow algorithm to comprehensively consider the static feature extraction based on face landmark and the optical flow difference feature. This, to a certain extent, makes up for the lack of static features. In comparison, there are still problems that need to be further studied and discussed.

① In the process of facial movement, the severity of the movement will affect the accuracy of key point marking and optical flow information feature extraction. The facial features extracted by facial paralysis patients with less effort or greater strength to complete the related facial movements have a certain degree of difference, which will affect the accuracy of facial paralysis grading assessment.

② This thesis uses the ETR algorithm to achieve feature point positioning, but the algorithm is not accurate enough to detect the feature points of patients with severe facial paralysis. Future work will mainly be improving the feature point positioning method to achieve accurate facial positioning of patients with severe facial paralysis.

③ Compared with the traditional method, this thesis has greatly improved the effectiveness of the model, but the grading performance of the suspected facial paralysis in the two levels of grade II and grade III is poor. Therefore, how to clearly distinguish between grade II facial paralysis and grade III facial paralysis has become the focus of the next step.

Reference

[1] Y. Jiang. Clinical Neurology [M]. Shanghai: Shanghai Medical University Press, 1999. (Chinese)

[2] M. Li, X. Huang, H. Xie. Psychological Effects on First-episode Bell's Palsy Patients and Their Mental Health Status [J]. China Journal of Health Psychology,2008(06):691-692. (Chinese)

[3] House JW, Brackmann DE. Facial nerve grading system[J].Otolaryngol Head Neck Surg, 1985, 93 (2) :146-147

[4] Yanagihara N. Grading of facial palsy [M]. Fisch U. Facial nerve surgery, proceedings of the Third International Symposium on Facial Nerve Surgery. Zurich:Kugler Med Publications,1976:337-343.

[5] Murty G E , Diver J P , Kelly P J , et al. The Nottingham System: Objective Assessment of Facial Nerve Function in the Clinic[J]. Otolaryngology -- Head and Neck Surgery, 1994.

[6] Ross B G , Fradet G , N Ed Zelski J M . Development of a sensitive clinical facial grading system[J]. Otolaryngology--head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery, 1996, 114(3):380.

[7] W. Yang, F. Wu, M. Zhang. Integrated Traditional Chinese and Western Medicine Evaluation and Efficacy Standards for Peripheral Facial Palsy (Draft) [J]. Chinese Journal of Integrative Medicine on Cardio-Cerebrovascular Disease, 2005, 003(009):786-787. (Chinese)

[8] VanSwearingen JM, Brach JS. The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. Phys Ther. 1996 Dec;76(12):1288-98; discussion 1298-300.

[9] Y. Li, Z. Gao, et al. Comparison of different facial nerve grading methods in the evaluation of peripheral facial nerve palsy[A]. Chinese Medical Association, Chinese Medical Association Otorhinolaryngology Head and Neck Surgery Branch. Chinese Medical Association 13th National Otorhinolaryngology-Head and Neck Surgery Academic Conference Paper Collection [C]. Chinese Medical Association, Chinese Medical Association Otolaryngology Head and Neck Surgery Branch: Chinese Medical Association,2013:1. (Chinese)

[10] I. Song, N. Y. Yen, J. Vong, J. Diederich and P. Yellowlees, Profiling bell's palsy based on House-Brackmann score [J]. 2013 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), 2013, pp. 1-6

[11] Y. Liu, Z. Xu, L. Ding, J. Jia and X. Wu. Automatic Assessment of Facial Paralysis Based on Facial Landmarks [J]. 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), 2021, pp. 162-167

[12] J. Ji, Y. Li, G. Li. Analysis of commonly used scales for evaluating facial paralysis [J]. Shanghai Journal of Acupuncture and Moxibustion,2009,28(07):421-422. (Chinese)

[13] Z. Guo et al. Deep assessment process: Objective assessment process for unilateral peripheral facial paralysis via deep convolutional neural network[J]. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 135-138.

[14] Z. Guo et al. An Unobtrusive Computerized Assessment Framework for Unilateral Peripheral Facial Paralysis [J]. IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 3, pp. 835-841, May 2018.

[15] X. Xiong and F. De la Torre. Supervised Descent Method and Its Applications to Face Alignment

[J]. 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 532-539

[16] Ngo T H, Seo M, Matsushiro N, et al. Quantitative analysis of facial paralysis based on filters of concentric modulation[C]. International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2016:1758-1763.

[17] Low D G. Object recognition from local scale-invariant features[J]. Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, pp. 1150-1157 vol.2.

[18] Low D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004.

[19] M. Ridha, W. Shehieb, P. Yacoub, K. Al-Balawneh and K. Arshad. Smart Prediction System for Facial Paralysis [J]. 2020 7th International Conference on Electrical and Electronics Engineering (ICEEE), Antalya, Turkey, 2020, pp. 321-327.

[20] G. J. Hsu and M. Chang. Deep Hybrid Network for Automatic Quantitative Analysis of Facial Paralysis [J]. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-7.

[21] S. Ji, W. Xu, M. Yang and K. Yu. 3D Convolutional Neural Networks for Human Action Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, vol. 35, no. 1, pp. 221-231,

[22] D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks [J]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 4489-4497.

[23] H. Li, J. Sun, Z. Xu and L. Chen. Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network [J]. IEEE Transactions on Multimedia, vol. 19, no. 12, pp. 2816-2831, Dec. 2017.

[24] B. Hasani and M. H. Mahoor. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 2278-2288.

[25] G. Storey, R. Jiang, S. Keogh, A. Bouridane and C. Li. 3DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework Using Fully 3D Convolutional Neural Networks [J]. IEEE Access, vol. 7, pp. 121655-121664, 2019, doi: 10.1109/ACCESS.2019.2937285.

[26] F. Wang, J. Cui, Z. Li. Research on Video Stabilization Algorithm Based on SIFT Feature Matching [J]. Information Security and Technology,2010(10):10-12

[27] Z. Li, S. Pundlik and G. Luo. Stabilization of Magnified Videos on a Mobile Device for Visually Impaired [J]. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 54-55

[28] G. Zhang, L. Yu and W. Wang. Video stabilization algorithm based on video object segmentation [J]. 2010 2nd International Conference on Future Computer and Communication, 2010, pp. V2-509-V2-512

[29] V. Avramelos, G. V. Wallendael and P. Lambert. Real-Time Low-Complexity Digital Video Stabilization in the Compressed Domain [J]. 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), 2018, pp. 1-5

[30] S. Wei, W. Xie, Z. He. Digital Video Stabilization Techniques ： A Survey [J]. Journal of Computer Research and Development,2017,54(09):2044-2058.

[31] Bleser G , Hendeby G . Using Optical Flow as Lightweight SLAM Alternative[C]. Science & Technology Proceedings, 8th IEEE International Symposium on Mixed and Augmented Reality

2009, ISMAR 2009, Orlando, Florida, USA, October 19-22, 2009. IEEE, 2009.

[32] Xia W. Design of Person Flow Counting and Monitoring System Based on Feature Point Extraction of Optical Flow [J]. 2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications, 2014, pp. 376-380

[33] Karayiannis N B, Tao G, Varughese B, et al. Discrete optical flow estimation methods and their application in the extraction of motion strength signals from video recordings of neonatal seizures[J]. IEEE, 2005.

[34] X. Wang, G. Zhang. Study on real-time detection method of moving target based on optical flow[J]. Computer Engineering and Applications, 2004, 40(001):43-46.

[35] Barron J L , Fleet D J , Beauchemin S S , et al. Performance of optical flow techniques[C] Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92. 1992 IEEE Computer Society Conference on. IEEE, 1992.

[36] C. Li, H. Bai, H. Guo, H. Liang. Moving object detection and tracking based on improved optical flow method[J]. Chinese Journal of Scientific Instrument,2018,39(05):249-256.

[37] Lucas B . An Iterative Image Registration Technique with an Application to Stereo Vision (DARPA)[J]. Proc Ijcai, 1981, 81(3):674-679.

[38] H. Xie, B. Yuan, W. Xie. Moving target detection algorithm based on LK optical flow and three-frame difference method [J]. Applied Science and Technology, 2016,43(03):23-27+33

[39] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform [J]. Proceedings of the IEEE, vol. 66, no. 1, pp. 51-83, Jan. 1978, doi: 10.1109/PROC.1978.10837.

[40] BRADSKI G, KAEHLER A.Learning open CV[M].Sebastopol,USA:O'Reilly Media,2009

[41] H. Liu, W. Meng and Z. Liu. Key frame extraction of online video based on optimized frame difference [J]. 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012, pp. 1238-1242, doi: 10.1109/FSKD.2012.6233777.

[42] A. Sarsenov and K. Latuta. Face Recognition Based on Facial Landmarks [J]. 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT), 2017, pp. 1-5

[43] J. Chen, V. Patel and R. Chellappa. Landmark-based fisher vector representation for video-based face verification [J]. 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2705-2709

[44] C. Sadu and P. K. Das. Swapping Face Images Based on Augmented Facial Landmarks and Its Detection [J]. 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020, pp. 456-461

[45] Kazemi V, Sullivan J. One Millisecond Face Alignment with an Ensemble of Regression Trees[C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2014.

[46] M. Yu, C. Zhang and M. Wu, "Research on Real-time Video Action Classification Based on Three-Dimensional Convolutional Neural Network," 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), 2018, pp. 339-343.

[47] H. Liu, F. Chang. Moving object detection by optical flow method based on adaptive weight coefficient[J]. Optics and Precision Engineering,2016,24(02):460-468

[48] Horn B K P, Schunck B G. Determining Optical Flow[C] Techniques and Applications of Image Understanding. International Society for Optics and Photonics, 1981.

[49] Bradski G R, Kaehler A. Learning OpenCV - computer vision with the OpenCV library: software that sees[M]. DBLP, 2008.

[50] S. Ding, B. Qi, H. Tan. An Overview on Theory and Algorithm of Support Vector Machines [J].

Journal of University of Electronic Science and Technology of China,2011,40(01):2-10

[51] Zhang S T , Wang F F , Fan D , et al. Research on the Majority Decision Algorithm based on Wechat sentiment classification[J]. Journal of Intelligent & Fuzzy Systems, 2018:1-10.

[52] Hsu C W , Lin C J . A Comparison of Methods for Multiclass Support Vector Machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):415-425.

[53] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[54] Modersohn L, Denzler J. Facial paresis index prediction by exploiting active appearance models for compact discriminative features[C] International Conference on Computer Vision Theory and Applications. INSTICC, 2016: 271-278

[55] He S, Soraghan J J, Oreilly B F, et al. Quantitative analysis of facial paralysis using local binary patterns in biomedical videos[J]. IEEE Transactions on Biomedical Engineering, 2009, 56(7): 1864-1870

[56] H. Ren. Quantitative assessment of facial paralysis based on content and spatiotemporal features[D]. Hunan: Central South University of Forestry and Technology, 2017

[57] M. Shen. Comprehensive evaluation method of facial paralysis classification based on optical flow difference characteristics[D]. Northwest University,2019

Research Results Obtained during the Degree Study

Paper to be published:

[1] **Wang B L**, Wu X M. Review of computer-based recognition and assessment of facial paralysis[J]. Chinese Journal of Medical Instrumentation Volume 46, No. 1 or 2, 2022

Acknowledgements

I would like to thank my tutor, Professor Wu Xiaomei. Professor Wu has profound professional knowledge, rigorous research attitude, innovative research ideas and meticulous attention to students. In the past two and a half years, Wu teacher not only give me guidance on the subject of the maze, but also care about my study life, I would like to express my gratitude to her.

Thanks to Teacher Xu Zhimin for her support. Teacher Xu has brought a lot of practical suggestions and opinions to my research direction, and also provided rich research data.

I would like to thank the students in the lab. It is a cherished experience to spend two years of postgraduate life with them.

Thanks to the students of Finland class, we have formed a profound friendship during the two and a half years of study together. I am very grateful to them for their help in my life, study and other aspects.

Thanks to my girlfriend, Li Danyu, who gave me a lot of support and encouragement in my study. We also created a lot of happy memories and will create more in the future.

Last but not least, I would like to thank my parents. Without their support, I would not be able to pursue my postgraduate study. They will enlighten me and share my happiness. What they have done is beyond measure. Thank you.