Disturbance Rejection with Compensation on Features

Xiaobo Hu[a], Jianbo Su[b,*], Jun Zhang[a]

[a]*Joint Institute of UM-SJTU, Shanghai Jiao Tong University, Shanghai*
[b]*Research Center of Intelligent Robotics, Shanghai Jiao Tong University, Shanghai*

Abstract

In pattern recognition tasks, the information from system input is modeled through a series of nonlinear operations, which include but not limited to feature extraction, regression, and classification. Both theoretically and practically, these operations are inevitably subject to internal modeling error and external disturbance, resulting at a performance challenge. Those state-of-the-art methods, e.g. Convolutional Neural Network and Transformer, still display significant instabilities and failures under practical applications, so comes a lack of generalization. Consequently, the more robust pattern recognition methods and related theories still merit a further study. This paper firstly reviews those state-of-the-art technologies in the field. The bottleneck of performances in those latest researches is associated with a lack of disturbance estimation and corresponding compensation. Therefore, the implications of disturbance rejection in pattern recognition field are further discussed from a control point of view. Then, the open problems are summarized. Ultimately, a discussion of the potential solutions, which is related to the application of compensation on features, is given to highlight the future study. Through the systematic review in this paper, the disturbance rejection in pattern recognition is developed into a control problem. Hopefully, more effective control technologies for the compensation on features can be used to improve the robustness of pattern recognition theoretically and practically.

*Corresponding author
Email addresses: xinaidd520@sjtu.edu.cn (Xiaobo Hu), jbsu@sjtu.edu.cn (Jianbo Su), zhangjun12@sjtu.edu.cn (Jun Zhang)

Keywords: Compensation, Disturbance rejection, Modeling error, Pattern recognition.

1. Introduction

As a branch of artificial intelligence, pattern recognition is designed to fulfill some manual heavy works [1]. However, there is an essential difference between artificial intelligence and human intelligence. Human makes qualitative analysis and judgments based on rational thinking. In contrast, pattern recognition is built with quantitative calculation. How to realize human intelligence and disturbance rejection is an important topic that so many scholars are still studying [2]. Generally, pattern recognition fulfills a mathematics modeling for those high-dimensional objects. This modeling can be realized through the manual feature extraction methods, such as SIFT [3] and Gabor [4], or the data driven ones, for examples, the widely used Convolution Neural Network (CNN) [5] and Transformer [6]. With the high-dimensional features, the regression problem in a localization task or the probability estimation in a classification task are accomplished. The working process of pattern recognition is shown in Fig.1.

The generalization of pattern recognition has always been a topic of importance [7]. Once there is a difference between the test sample and the training one, the success rate always decreases in varying degrees, for examples, the cross age [8] and cross races facial recognition [9]. This is a bottleneck in the wider promotion of pattern recognition. In the recognition process, modeling errors are unavoidable, e.g. the depth of neural network, the operation of convolution or pooling, etc. Moreover, the input data is often coupled with various disturbances. For instance, there are several prominent factors that lower the success rate in facial recognition, which include but not limited to occlusion, age and gender [10]. Except the disturbances in facial recognition, some other disturbances include the noise problem in speech recognition [11], focusing problem in image recognition [12], object appendage in gait recognition [13], and pedestrian disturbance in road lane detection [14]. All these representative disturbances

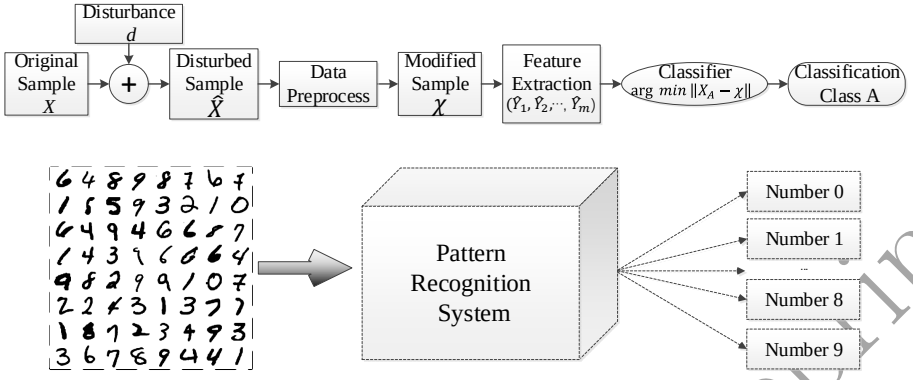

Figure 1: A typical pattern recognition process.

widely exist. In a word, it is almost impossible for any recognition system to completely get rid of its internal *modeling error* and avoid external *disturbance*[1].

Researchers have made many successful attempts to improve the generalization. In facial recognition, there are diverse methods to improve the quality of input image, so as to improve the performance, e.g. multimodal information complementation [16] and image recognition under near-infrared light [17]. These methods are mainly designed with the digital signal processing which focuses on noise or outlier [18]. Furthermore, the facial features in [8] are decomposed into the orthogonal age-related features and identity-related ones, so as to improve the efficiency for a cross age facial recognition. In [19], latent space projection is applied into the class imbalance problem, which is generated by the inconsistent image resolution. In addition to the applications in facial recognition, there are also some other efficient methods. In [20], the Mel Frequency Cepstral Coefficients (MFCCs) feature is used to separate the current speaker's voice from noise, thus improving the quality of speech recognition. Besides, with the rapid development of artificial intelligence, many deep learn-

[1]The definition of *modeling error* and *external disturbance* are given from a system control point of view in this paper. In some existing researches for pattern recognition, they may be defined as *Aleatoric* and *Epistemic* uncertainty respectively [15].

ing based techniques have been realized into the pattern recognition field, which show a strong robustness [21].

However, there is no uniform disturbance rejection criterion due to the diversity of different tasks. The modeling error is under cover by the deeper and deeper networks [6], which usually fulfill the objective tasks with a black-box mapping. Besides, the larger and larger datasets are used to approximately cover all kinds of external disturbances, including the image noise, identity characteristics, and cross-domain difference [22]. These open-loop principles usually ignore the modeling and compensation of internal and external disturbances. All of the disturbances are regarded as an inconsequential factor that needs to be avoided or eliminated. However, the depth of neural networks, the size of datasets, and the VRAM of GPU can not be increased endlessly. Therefore, a novel solution is taken by the close-loop principle of *feedback*, in which a compensation point of view is used to improve the recognition performances [23].

Compensation is a core principle in the control science field [24]. In order to resist the internal and external disturbances in electromechanical systems, researchers have proposed different kinds of control theory, such as PID [25], Active Disturbance Rejection Control (ADRC) [26], Disturbance Observer Based (DOB) control [27], and Sliding Mode Control (SMC) [28]. The core principle of these theories is to construct a feedback loop based on the error between expected output and practical output, so as to realize a *disturbance rejection control*. Currently, the feedback in pattern recognition is mainly used as an information compensation [23]. In this paper, the compensation in pattern recognition is further studied within a framework of disturbance rejection.

The modeling error and the external disturbance are two main factors affecting the overall performance of a controllable system. However, the academic research and industrial application still lack a full-developed theoretical analysis for the modeling error and external disturbance in a recognition system. Although the denoise technology has been widely developed in pattern recognition field, the more systematic and comprehensive development from a disturbance

rejection control point of view is still an open problem. Therefore, this paper will firstly summarize the modeling error and external disturbance in those state-of-the-art pattern recognition models, demonstrated in Section 2. Then the open problems are developed in Section 3. Furthermore, the disturbance rejection problem is studied in a control point of view in Section 4, in which the potential solutions and future study are discussed. Ultimately, the conclusions are drawn in Section 5.

2. An Overview of Pattern Recognition: State-of-the-art

As shown in Fig.1, a pattern recognition task is usually realized with three major operations: (1) data preprocess, (2) feature extraction, and (3) classification or localization. This subsection summarizes some mainstream theories of pattern recognition in these operations. Then, the disturbance sources will be analyzed for a further shortcoming concentration on these state-of-the-art technologies.

2.1. Working Flow of Pattern Recognition

2.1.1. Data Preprocess

Data preprocess is one of the main techniques for improving recognition performance. For example, the smoothing filter or the Gaussian filter is widely used in image denoising [29]. More in-depth, with the research of adversarial learning, it has been applied into image restoration and appendage elimination [30]. Generally speaking, the designer needs to choose a suitable method for data preprocess through a prior evaluation of the application scenarios. However, with the expanding scenarios, the lack of robustness for traditional data preprocess often results in a performance degradation [31]. When the new application brings a new kind of disturbance, the system needs to be redesigned.

2.1.2. Feature Extraction

In addition to data preprocess, the more robust feature extraction is another important mean to improve the performance. Some manual extraction methods,

such as SIFT [3] and HARR [32], mainly consider the grayscale information between pixels inside an image. Besides, through a large number of well-annotated training samples [33], the system designed with machine learning can learn how to interact with the external environment then complete some specific tasks in the high-dimensional state space [7]. As a representative machine learning technology, deep learning is designed to simulate the structure of human brain neurons [34]. When the high-dimensional data is transferred into a deep neural network, the relationship between input and output is established with a layer by layer structure. The internal parameters of each neuron are obtained by a performance optimization with the loss function [35].

2.1.3. Classification or Localization

Following feature extraction, the ultimate classification or localization subtask is usually completed by an activation function, such as Softmax or Sigmoid [36]. Some current works about disturbance rejection in the field are still focused on the more robust classifier [37]. The more complicated if those application scenarios are, the more contaminated input data will make the difference and similarity between the recognizable objects become blurry. How to minimize the distance between similar objects and maximize the distance between heterogeneous ones simultaneously is still an important problem [38]. The output error in a recognition system is reflected by an increase on uncertainty, or even a wrong recognition result. With some current failure cases, such as the recognition errors between facial objects and the traffic accidents caused by automatic driving [39], pattern recognition still has a long distance to become a full developed *artificial intelligence*.

2.2. System Disturbances

When a system works, disturbance always exists. Since the birth of pattern recognition, the concern about its robustness has never stopped [40]. This subsection analyzes the generation of disturbances in a pattern recognition task.

2.2.1. Disturbance in Data

Different kinds of disturbances are always coupled in the input data, so comes the Aleatoric uncertainty [15]. This part of disturbance mainly comes from the external noise outside the recognition system, namely an *external disturbance*. A typical case is the success rate reduction in cross dataset validation. When some current systems are assessed with a cross dataset validation, most of them fails to retain a good performance, which owes to the variation of superimposed disturbances contained in different datasets, e.g. the different races and cultures in facial recognition [9]. Different kinds of dataset composition will have a great difference on the success rate, no matter in a training stage or a testing one [41]. To obtain an effective data preprocess, it is necessary to build a prior knowledge towards the previous samples, as mentioned in Section 2.1.1. However, there is still limited metric to judge an effectiveness of specific data preprocess operation. Moreover, when new data is received, most of the systems have entered an *autonomous* state at this moment. Therefore, the data preprocess needs to be taken without human revision. Once those disturbances contained in the input are changed, such as race variation for human face, road condition variation for automatic driving, and background noise variation in speech segment, the original data preprocess may become far away from an optimal solution. The useful information may be removed, while some useless one is kept. The current solution is a manual intervene, with which the preset process or the system parameters are adjusted.

2.2.2. Disturbance in Feature Extraction

The main sources of disturbance in feature extraction include but not limited to: the gradient estimation, the selection of kernel space, the setting of hyper-parameters, etc. The internal parameters of a recognition model are designed with human prior knowledge or a performance optimization. However, due to the contamination in training dataset and the incomplete system description, such a design process leads to the Epistemic uncertainty [15]. As a result, the *internal modeling error* always exists in a pattern recognition system. For ex-

ample, the complete modeling in independent component analysis (ICA) [9] is regarded as a feature space consisting of a group of independent vectors, e.g. $(Y_1, Y_2, \cdots, Y_M)$. Nevertheless, it is usually unavailable. The estimated feature space is built as $(Y_1, Y_2, \cdots, Y_m)$. The estimated dimension m is smaller than the ideal M. Then, the Input X is represented by:

$$X = \sum_{j=1}^{M} \alpha_j \cdot Y_j = \sum_{j=1}^{m} \alpha_j \cdot Y_j + \sum_{k=1}^{M-m} \alpha_k \cdot Y_k \approx \sum_{j=1}^{m} \alpha_j \cdot Y_j \ , \tag{1}$$

in which α_j is the projection coordinate. Hence, the error convergent condition is:

$$sup \ || \sum_{k=1}^{M-m} \alpha_k \cdot Y_k || \leq \varepsilon \ . \tag{2}$$

The upper limit ε should be bounded. Except the incomplete dimension, there is still uncertainty on the loss function definition [42]. Moreover, the orthogonality between Y_j still needs to be further optimized by knowledge distillation [43]. All these factors lead to a cumulative modeling error.

2.2.3. Failure of Recognition

Due to the mentioned external disturbance in data, and the modeling error in feature extraction, a recognition task may fail, e.g. the defeated models under Adversarial Attack [44], the poor-performed deep networks in engineering applications [45], etc. The ultimate goal for pattern recognition is to distinguish some input objects and output their corresponding information. The system output can be obtained by different kinds of classifier, which are object-oriented for some given classes, or by a regression of position coordinates. For example, there are intraclass and interclass relationships in a classification task [46]. The recognition failure is given by a mistaken calculation of distance between and within clusters. In Fig.2, Class A and B, corresponding to *Eskimo husky* and *Shetland sheepdog* are separated by $\mathcal{A}Y_1 = Y_2$ in the two-dimensional feature space. However, a soiled *Eskimo husky* is misclassified to be *Shetland sheepdog* by the pretrained VGG-16.

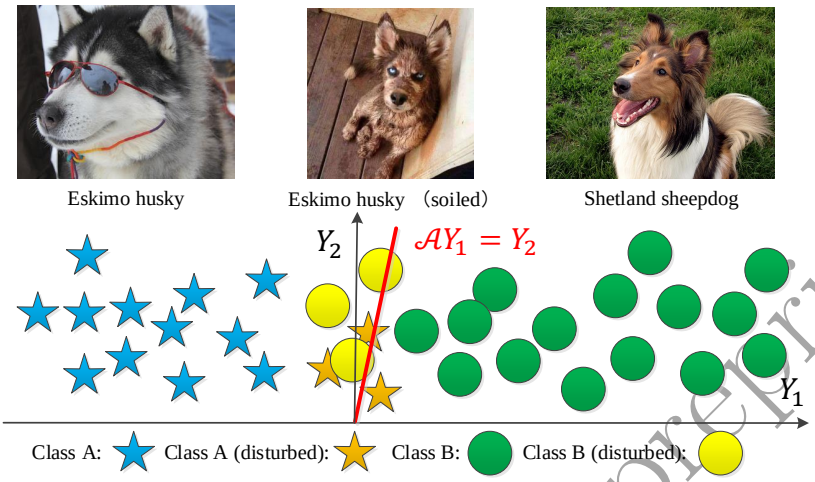

Figure 2: An example of failure on CNN. The tested CNN is VGG-16 pretrained on ImageNet.

2.3. Current Technologies on Disturbance Rejection

The development of pattern recognition field has been focused on the disturbance rejection problem, no matter for the far past development of digital image processing, or those hot topics on artificial intelligence, e.g. the latest ChatGPT [47, 48, 49]. The *Big Data* and *Big Model* become main principles in the field. For example, the classification system based on machine learning is designed through the maximization for interclass difference as well as intraclass similarity towards a training dataset. Some state-of-the-art disturbance rejection techniques are reviewed in this subsection.

2.3.1. Data Augmentation

In order to make the training dataset closer to the real scenario, researchers use some controllable disturbances to modify those clean training datasets into contaminated ones [50]. In [51], the random erasure blocks some pixels in the input image, so as to train a more robust system. Besides, a fuzzy dataset in [52] is generated by the clear still images. Then, the augmented dataset is used to make up for the deficiency of real video. Moreover, the widely used Generative Adversarial Networks [53] are also applied into the data augmentation work [54].

In summary, these methods are implemented in the designing and training stage. From a system control point of view, once a disturbed signal is input, the system will obtain a corresponding response under its internal working mechanism. This response reflects the input-output relationship under a disturbance excitation. Such information can be used as the feedback signal to construct a closed-loop system, so as to improve the robustness [55]. However, there is still a lack of systematic design method and mathematical basis to realize the concept of *disturbance feedback* in pattern recognition.

2.3.2. Disturbance Representation

Different disturbance sources must bring variational impacts on diverse systems [21]. An effective way to improve performance is to identify the types of disturbance with an explicit representation. Then a distinction of the disturbance-related and disturbance-unrelated information needs to be done. In [8], the age-related representation of facial image is modeled by sphere polar coordinates, which are orthogonal to each other. Similarly, the extended dictionary learning in [56] is used to express the disturbance of appendages and clothing in gait recognition.

The advantage of disturbance representation is to expand the feature expression of an input object, so as to identify and compensate for the disturbance. In contrast, the disadvantage is that it must be expressed for a specific disturbance. When there are different kinds of disturbance, even for the unknown ones, the dimension of feature space gets bigger and bigger.

2.3.3. Robust Features

To guarantee a robust recognition, one of the necessary conditions is the appropriate features. In ICA based expression recognition task, the race-related features and expression-related features are represented by two groups of mutually independent vectors, then the race-related features can be selected through a maximization of mutual information [9]. Besides, the influences of different disturbances are varied for different feature extraction methods. According to

the variation generated by specific disturbances, the separability of those extracted features can be evaluated, such as the Fisher discriminant ratio [57]. If the Fisher discriminant ratio of a recognition system gets bigger, there are more features could distinguish different classes in the objective task. Besides, the state-of-the-art technologies, e.g. the deeper and deeper network [58], the Attention mechanism [6], the cross compensation between features [59], and so forth, all point to the more robust features.

The core principle for this kind of method is to select the extracted features which are more robust towards some given disturbances, no matter these disturbances exist in the collected dataset or in the manual augmented one. However, the disturbances need to be specified with a prior knowledge. So, it is difficult to be generalized into some more complex situations, especially those cases containing several kinds of unknown disturbance sources [60].

3. Open Problems

The topic of disturbance rejection comes from the control science field [26], which usually consists of three necessary steps: (1) The disturbances are estimated, (2) The controllable input is constructed with a feedback gain, and, (3) The compensated system completes the objective task. The compensation principle in disturbance rejection has been applied into diverse realizations in pattern recognition field. However, the more systematical design of disturbance rejection loop in a recognition system remains understudied. Those state-of-the-art researches in pattern recognition are mainly taken on a model level rather than a system level. Most of the latest researches focus on the improvements of recognition model, e.g. the deeper and deeper networks or bigger and bigger training dataset. A systematic error compensation is rarely considered. Therefore, the disturbance rejection in pattern recognition is designed passively but not actively. To bridge the pattern recognition field and control science field, the open problems towards disturbance rejection in pattern recognition are summarized from a system control point of view in this section.

3.1. Evaluation Metrics

In system control field, the disturbance rejection ability of a system is usually evaluated under the band-width, response time, peak value, etc [61]. In [62], the definition of disturbance coefficient is given to evaluate the robustness of a closed-loop system. However, the operation of those existing methods in pattern recognition still needs more evaluation proof to explain its robustness [63]. The evaluation metrics is one of the determination factors affecting a system design. Usually, the tracking problem in pattern recognition tasks is a minimum error track on the high-dimensional object, such as the clustering with Center Loss [64]. The current widely used robustness evaluation is the success rate or the related metrics on some target datasets [1, 14, 65]. From a mathematical point of view, these success rate metrics are usually described by a one-dimensional real number belonging to the interval of [0,1]. How to evaluate the complex operations in a recognition system merits further study.

3.2. Features

Feature extraction is always the core issue in pattern recognition [66]. No matter for the useful information, or the rejected disturbance, all of them need to be represented under a complete enough feature space. The controllable feedback for unknown disturbance and modeling error needs to be made towards the compensated recognition system. The complete description for those disturbances in a recognition task merits a further study. What kind of noise will lower the system performance? What kind of disturbance can be handled by the compensation mechanism? These problems need to be further analyzed in the specific recognition system.

3.3. Disturbance Observation and Compensation

The observation and compensation for disturbances needs to be constructed with the *output error*, e.g. the interclass similarity and intraclass difference in a

classification task. In the case of Multi-Layer Perceptron, the intermediate i^{th} operation is usually taken as follows:

$$\begin{cases} \widehat{Y}_{i+1} = \sigma(\widehat{Y}_i \cdot \omega_i + b_i) \\ Y^*_{i+1} = \sigma(Y^*_i \cdot \omega_i + b_i) \end{cases}, \tag{3}$$

in which $\widehat{Y}_i$ and Y^*_i correspond to the disturbed and ideal feature respectively. Usually, Y^*_i is defined to be a Cluster Center [64]. The weight matrix, the bias, and the activation in a layer-wise calculation are represented by ω, b and $\sigma(\cdot)$. Then, the disturbance observation $\hat{d}$ is constructed by the given information, e.g.

$$\hat{d} = \mathcal{F}_O\left(X, \widehat{Y}_i, Y^*_i\right) . \tag{4}$$

The observation $\mathcal{F}_O\left(X, \widehat{Y}_i, Y^*_i\right)$ is realized with a Bidirectional Matrix Feature Pyramid Network in [59], and a Semantic Feature Pyramid Network in [67]. Then, the object detection system, designed with a Feature Pyramid Network (FPN) framework, is compensated, e.g. $\widehat{Y}_{i+1} = \sigma(\widehat{Y}_i \cdot \omega_i + b_i + \hat{d})$ in [23]. However, the prior knowledge about Y^*_i maybe unavailable, especially in an Unsupervised Learning case [68]. The more realizations of $\mathcal{F}_O(\cdot)$ remains understudied.

3.4. Forward and Reverse Process

Disturbance rejection needs to eliminate the noise effect from a disturbed input. Besides, the intermediate modeling error also should be compensated. These observation and compensation works are realized towards the total disturbance by ESO in [26], DOB in [27], BMFPN in [59], and SFPN in [67]. The disturbance observation forms a reverse process comparing with the forward one in a recognition task, so comes a *close loop*. However, the nonlinear operations, e.g. convolution and pooling, leave an open problem into the inverse model. In GANs [53], the structure of Generator $\mathcal{G}$ and Discriminator $\mathcal{D}$ are usually different. When and how the recognition model can be inverted merits a further study [69]. If the model can not be reversed theoretically, the disturbance observation and its further compensation would become a more challenge problem.

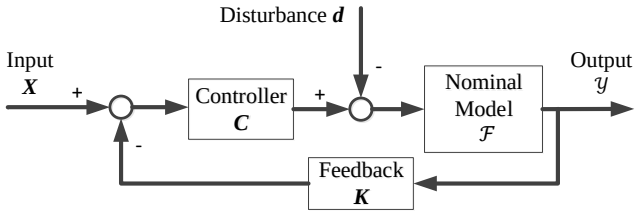

Figure 3: A close-loop controlled system with feedback.

4. Solution: Disturbance Rejection Control in Pattern Recognition

In control theory, the disturbance information is used to construct a closed-loop feedback, with which an improvement on the compensated system can be achieved [70]. However, the pattern recognition model is quite different from the establishment of state equation in some typical controlled systems. Moreover, the input in pattern recognition is more complex. Luckily, some successful trials have been realized, e.g. the Feature Compensation in FPN [23], the Two-stream Network in visual detection [71], etc. This section will summarize the realization of disturbance rejection control theory in pattern recognition.

4.1. Disturbance Rejection Control

4.1.1. Classical Controlled Model

A typical closed-loop controlled system is shown in Fig.3. The Input X, coupled with Disturbance d, is processed. Then the Output $\mathcal{Y}$ is obtained. The modeling error of Nominal Model $\mathcal{F}$ leads to an error between the expected output and the practical one. Such errors are stabilized by the Controller $\mathbf{C}$ and the Feedback Loop $\mathbf{K}$. The m^{th} order system is formulated as follows:

$$\begin{cases} \dot{Y}_1 = Y_2 \\ \quad \cdots \\ \dot{Y}_{m-1} = Y_m \\ \dot{Y}_m = d(t) + F(Y,t) + B(Y,t) \cdot U(t) \\ \mathcal{Y} = Y_1(t) \end{cases}, \tag{5}$$

in which the time-related external disturbance is marked with $d(t)$. The controllable input is given by $U(t)$. The modeling error is generally reflected in the construction of $F(Y,t)$ and $B(Y,t)$. Besides, the approximation of higher order terms or the unknown nonlinear functions in $F(Y,t)$ and $B(Y,t)$ is often ignored, which also contributes to the modeling error. A controllable system is required to output the expected X under an error convergence, i.e. the Output $\mathcal{Y}$ needs to be approximately equal to X.

4.1.2. Disturbance Observation

The core problem of disturbance rejection in a controllable system is how to estimate the disturbance then compensate for the error [72]. The observation methods are varied based on the practical systems and the control strategies.

Time Domain Sense. In ADRC, the combination of Tracking Differentiator (TD), Extended State Observer (ESO), and Nonlinear State Error Feedback (NLSEF), is used to compensate for the total disturbance [26]. The disturbance in the m^{th} order state variables, i.e. $\dot{Y}_m = d(t) + F(Y,t) + B(Y,t) \cdot U(t)$, is determined by the modeling error, real-time input and external disturbance. The principle of ADRC is to take these disturbances as a total disturbance. Then the total disturbance is compensated with $\dot{Y}_{m+1}$ in a reconstructed system, e.g.

$$\begin{cases} \dot{Y}_1 = Y_2 \\ \quad \cdots \\ \dot{Y}_{m-1} = Y_m \\ \dot{Y}_m = Y_{m+1}(Y,t,U(t)) + \bar{B}(t)U(t) \\ \mathcal{Y} = Y_1(t) \end{cases} . \tag{6}$$

The additional state variable is $Y_{m+1}(Y,t,U(t))$, so comes the name with *Extended State Observer* (ESO), i.e.

$$Y_{m+1}(Y,t,U(t)) = d(t) + F(Y,t) + (B(Y,t) - \bar{B}(t))U(t) \ , \tag{7}$$

in which $\bar{B}(t)$ is an approximation of $B(t)$. Through an expansion of the State $Y_{m+1}(Y, t, U(t))$, the total disturbance can be estimated and moreover compensated. The key element is to use the difference between practical output and expected output for a disturbance observation. Then the compensation is applied, so that the error is stabilized.

Laplace Domain Sense. In Laplace domain, the input-output relationship can be formulated by a transfer function. A representative disturbance rejection control method in Laplace domain is the Disturbance Observer Based (DOB) strategy [73]. Regarding to the DOB framework, the disturbance observation is expressed as follows:

$$\hat{d} = \mathcal{F}^{-1}(s) \cdot (\mathcal{F}(s) \cdot (U + d) + \zeta) - U \ , \tag{8}$$

in which ζ is the observation noise. Then, the system output is given by:

$$\mathcal{Y} = \mathcal{F}(s) \cdot \left(X - \hat{d} + d\right) \approx \mathcal{F}(s) \cdot X \ . \tag{9}$$

Towards some industrial application, it may be difficult to represent the Nominal Model $\mathcal{F}$ by a group of explicit function, which is similar to the complicated neural network. Moreover, the inverse model of $\mathcal{F}^{-1}$ may be unattainable with specific hardware equipment. Usually, this problem is solved by the design of Q *filter* [74]. All these characteristics of DOB fit the ones in pattern recognition, especially for the open problem of reverse process in Section 3.4.

Pattern Recognition Sense. The disturbance d in pattern recognition, summarized in Section 2.2.1, may be generated by a single source or multiple sources, e.g. the rotation angle and the handwriting habits in Fig.1 may affect the recognition result individually or jointly. Besides, these sources can be mutual independent or dependent. Moreover, the intermedia modeling errors, summarized in Section 2.2.2 always exist. Referring to Section 3.3, the designing principle of disturbance observation in pattern recognition is still an open problem. Firstly, the external disturbance d varies according to different application scenarios.

Moreover, the observation methods still lack a strict mathematical foundation. For example, it is still difficult to observe the disturbance of soil in Fig.2. Although there have been some explorations in Section 2.3.2, how to observe d from those intermedia errors in a recognition process remains understudied.

4.1.3. Controllable Input

A natural problem following disturbance observation is the construction of feedback, i.e. error compensation. With an error between the practical output and the expected one, the disturbance observation is obtained, so that a further compensation can be given towards the controllable input. The compensation is used to form a feedback loop in the controllable system. Then the whole system becomes a close loop. In control theory, **feedback** is a core principle to ensure robustness [24]. Nowadays, the most widely used classical control method in the industry must be PID [75]. The control signal of PID is constructed by:

$$U(t) = k_p \cdot e + k_i \cdot \int_0^t e d\tau + k_d \cdot \frac{de}{dt} \ , \tag{10}$$

in which k_p, k_i, and k_d are pre-set parameters. Error e is the difference between the tracked Input X and the real-time Output $\mathcal{Y}$. An adequate design of PID can realize an unbiased tracking for Input X asymptotically [76].

4.2. Control on Pattern Recognition

The modeling in a recognition system is quite different from the one in traditional controllable system. Therefore, the specific compensation should be designed based on the high-dimension feature space in pattern recognition.

4.2.1. Hierarchical Controlled Model of Pattern Recognition

The input of a recognition system is usually a high-dimensional object, e.g. the signal of X to be tracked in Fig.3 is an image or a segment of voice. Besides, the output of a pattern recognition system is an identification or localization

for the input [36]. Hence, the pattern recognition task with Input X under a m-steps feature extraction is modeled as follows:

$$\begin{cases} \widehat{X} = \mathcal{F}_D(X, d) \\ \chi = F_0(\widehat{X}) \\ \widehat{Y}_1 = F_1(\chi) \\ \widehat{Y}_2 = F_2(\widehat{Y}_1) \\ \quad \cdots \\ \widehat{Y}_m = F_m(\widehat{Y}_{m-1}) \\ \widehat{\mathcal{P}}_{\mathcal{N}} = F_{\mathcal{N}}(\widehat{Y}_m) \end{cases}, \tag{11}$$

in which $\mathcal{F}_D(\cdot)$ is a data acquisition function for error-free Sample X, corrupted by the external Disturbance d. The acquired $\widehat{X}$ can be regarded as an image or a segment of voice. The *modeling, disturbance*, and *error* in this hierarchical model is demonstrated as follows:

- The disturbance d may be a variation of observation angle [77] or an occlusion in facial image [78]. In the current researches, this kind of disturbance is regarded as an uncertain noise embedded in the Input X. Usually, it is named with *Noise* in the literatures. In this paper, this kind of *noise* is classified as external disturbance in a controlled recognition system, which is independent with the other disturbance source of modeling error.
- The elimination of d should be taken by a data preprocess $F_0(\cdot)$. The realization of $F_0(\cdot)$ has been studied by voice denoise in [66], pedestrian detection in [79], etc. Then a denoise Sample χ is obtained.
- The feature extraction of $F_i(\cdot)\,, i = 1, \cdots, m$ can be realized by the widely used convolution, pooling, or full-connected layers [7]. After that, the recognition process of $\widehat{\mathcal{P}}_{\mathcal{N}}$ is obtained by an activation function $F_{\mathcal{N}}(\cdot)$ with the deepest Feature $\widehat{Y}_m$.
- The final recognition result is given by the respective information corresponding to $\widehat{\mathcal{P}}_{\mathcal{N}}$, e.g. $\mathcal{N}$ is the amount of objective classes in a classification

task. The realization of $\widehat{\mathcal{P}}_{\mathcal{N}}$ can be a closest neighbor search or a probability estimation [7]. For example, a conditional probability estimation of $P_A = P(X \in Class\ A) = P(A|\widehat{Y}_1, \widehat{Y}_2, \cdots, \widehat{Y}_m)$ is obtained with all these extracted features.

- The modeling errors in this recognition model include: (i) it is difficult to guarantee that all these m steps of feature extraction is robust enough to fulfill a recognition task, (ii) the feature space can not be validated to be an inner product space, and, (iii) the dimension of feature space can not be determined theoretically.

- **Example**: In MNIST recognition task, disturbance d may be generated by the rotation or occlusion of an image. Then, Input X is one of the basic decimal digits of $0 \sim 9$. Due to the variation of d in application scenarios, the obtained features of $\widehat{Y}_i$, $i = 1, \cdots, m$ may not be able to represent all of the information of X, so that the recognition result may be wrong.

4.2.2. Error System

When disturbance exists, the error system corresponding to Equ.(11) is modified as follows:

$$\begin{cases} e_0 = X - \chi \\ e_1 = Y_1^* - \widehat{Y}_1 \\ e_2 = Y_2^* - \widehat{Y}_2 \\ \quad ... \\ e_m = Y_m^* - \widehat{Y}_m \\ e_{\mathcal{N}} = \mathcal{P}_{\mathcal{N}}^* - \widehat{\mathcal{P}}_{\mathcal{N}} \end{cases}, \tag{12}$$

in which the ideal recognition $\langle X, Y_1^*, Y_2^*, \cdots, Y_{m-1}^*, Y_m^*, \mathcal{P}_{\mathcal{N}}^* \rangle$ is used to obtained all intermedia errors. The error system (12) can be separated into three parts. Firstly, e_0 gives an error in the input data acquisition, which belongs to the disturbance of external noise, discussed in Section 2.2.1. Moreover, the error

of e_i $(i = 1, \cdots, m)$ is generated by the feature extraction $F_i(\cdot)$, discussed in Section 2.2.2. When these internal and external disturbances act on the final result of $\widehat{\mathcal{P}}_{\mathcal{N}}$, there always be a recognition error $e_{\mathcal{N}}$ compared with the desirable result $\mathcal{P}^*_{\mathcal{N}}$, which is related to the failure in classification, demonstrated in Section 2.2.3.

4.2.3. Error Convergence Condition

The error convergence is one of the most important part in the system control field. In [80, 81, 82], additional proofs are given about the theoretical performance of the disturbance rejection control. However, it is still a challenge problem in almost all of the fields in pattern recognition [83]. Regarding to the controlled system in Equ.(5), the ultimate goal for error stabilization is to realize a track on the output, i.e. $||\mathcal{Y} - X||_P \leq \varepsilon$ in a L_P norm. The similar form in an ideal recognition should be the track on $\mathcal{P}^*$ with $||\widehat{\mathcal{P}} - \mathcal{P}^*||_F \leq \varepsilon$. The measurement criterion F includes the Euclidean distance, Wasserstein distance, etc [84].

A correct recognition only needs to ensure that the minimum distance between an input sample and its corresponding class is still the objective one, rather than the others, which leaves a loose convergence. For example, a binary classification between *Class A* and *Class B* is donated by $A = (Y_A^1, Y_A^2, \cdots, Y_A^m)$ and $B = (Y_B^1, Y_B^2, \cdots, Y_B^m)$ in a m-dimensional feature space. Then, a *Class A* sample is observed by $\hat{A} = (Y_{\hat{A}}^1, Y_{\hat{A}}^2, \cdots Y_{\hat{A}}^m)$. The mis-recognition occurs if a wrong closest condition is satisfied, e.g.

$$||\hat{A} - A||_F > ||\hat{A} - B||_F \ . \tag{13}$$

How to avoid Equ.(13), and furthermore stabilize error system (12), is the fundamental problem in a stability proof. The stability condition for some other recognition form, e.g. a localization, is similar to this classification one. A similar deduction is omitted.

Table 1: Comparison of Nomenclature between Control Science and Pattern Recognition.

Nomenclature	Control Science	Pattern Recognition
System modeling	Dynamic system, e.g. ODE, PDE	Hierarchical process
System input: X	Controllable variable	Image, voice, etc.
System output: $\mathcal{Y}$	Controllable object	Classification or localization
Tracking object	Equilibrium	Cluster or localization center
Disturbance (Internal)	Nonholonomic or nonlinear constraint	Network depth, feature completeness, etc.
Disturbance (External)	Noise $d(t)$	Image noise, voice background, etc.
Description operator	State variable: $\dot{Y}_1, \cdots, \dot{Y}_m$	Feature: $Y_1, \cdots, Y_m$
Operation	Differential, integral, and difference	Convolution, pooling, and linear regression
SOTA	PID, ADRC, DOB, etc.	CNN, Transformer, etc.
Feedback design	ESO, DOB, NLSEF, etc.	Feature compensation, attention, etc.
Parameter turning	Manual adjustment	Loss optimization
Stability condition	Error convergence	

4.2.4. From Model to System

Currently, the state-of-the-art researches in pattern recognition are mainly focused on the more robust models. The disturbance rejection development is still stuck at the model level, not yet elevated to a systematic control level. The control method on pattern recognition is quite different from the ones in classical control science field. A solid foundation from the mathematics point of view merits a further study on the recognition process. A comparison between the control problem in a classical dynamic system and the one in a pattern recognition process are summarized in Table 1. The system model of pattern recognition is built with Equ.(11) in a hierarchical computation form. However, most of the latest researches is taken based on the differential or difference dynamical system, e.g. the one in Equ.(5). Besides, the tracking object in a traditional control system is *Equilibrium*. In contrast, the tracking object in a pattern recognition task should be a cluster or localization center. Furthermore, the key element in a dynamical system is state, i.e. $\dot{Y}$ in Equ.(5), while the one in a pattern recognition system is feature, i.e. $\widehat{Y}$ in Equ.(11). All these differences deserve a further study in not only the control science field but also the pattern recognition field. Pattern recognition has been developed for decades. Many

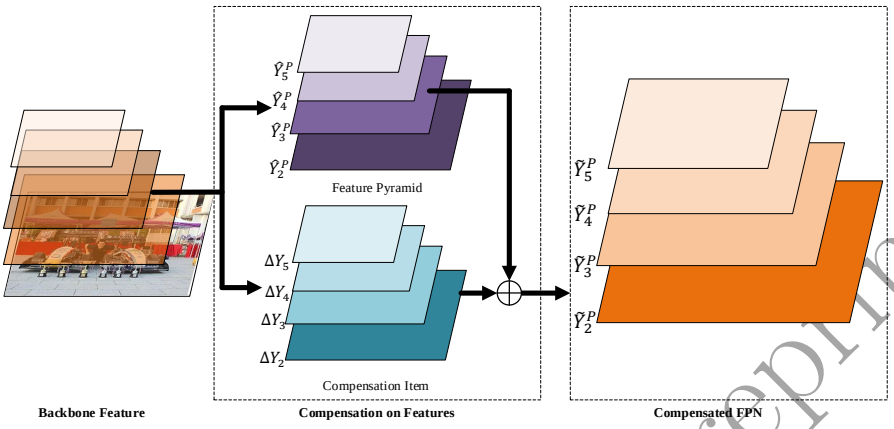

Figure 4: A typical compensation on features (Example: FPN) [23].

theories and methods have been proved to be effective and robust in practice. However, how such a mechanism could be explained for its rationality is an important issue that so many scholars still question on [2].

4.3. Realization of Feedback: Compensation on Features

The state-of-the-art researches of pattern recognition focus on two major principles, i.e. (1) Data is all you need; and (2) GPU is all you need. If the collected training dataset is closer to the application scenario, the model will perform better. Besides, if more GPU devices are used to train and test a recognition model, the recognition performance will also be better. All of these principles point to the robustness of extracted features. Therefore, most of the feedbacks in pattern recognition is realized on features. For example, a representative application in Feature Pyramid Network (FPN) is shown in Fig.4. The compensation for Feature $\widehat{Y}$ is taken as follows:

$$\widetilde{Y} = \widehat{Y} + \Delta Y \ , \tag{14}$$

in which the compensation item of ΔY fulfills a disturbance rejection. Then, the fused Feature $\widetilde{Y}$ should improve the recognition robustness. Several representative cases about how to construct ΔY are summarized in this subsection.

4.3.1. Information Compensation in FPN

The dilution disturbance in Feature Pyramid Network is associated with an inevitable contradiction between semantic and localization information [85]. The deeper and deeper network is used to obtain more semantic information for the input in an object detection tasks [5]. However, the less localization information is reserved in the deepest features. The compensation mechanism for those hierarchy features is used to solve this problem. The compensation item of ΔY is realized with a Bidirectional Matrix in [59]. Besides, it is realized with a Semantic Feature Pyramid in [67]. The compensation on feature is constructed to make up for the semantic information in those low level features, e.g. $\widehat{Y}_1$ or $\widehat{Y}_1$ in Equ.(11). Furthermore, the feature compensation makes up for the localization information in those high level features, e.g. $\widehat{Y}_{m-1}$ or $\widehat{Y}_m$.

4.3.2. Two-stream Ideology

The static image-based framework is not robust enough to reject those internal and external disturbances in a video-based recognition task [71]. On one hand, the frame by frame operation needs a heavier computation cost. On the other hand, the information of variations in the input sequence also contributes to the objective tasks. For example, in Parkinson detection, the video can show abnormal expressions of a patient, which can not be obtained from a static image [86]. Therefore, the compensation mechanism in Two-Stream network is used to obtain a more robust recognition system. Specifically, the spatial network, e.g. a Convolutional Neural Network, extracts those identity and static features of the input, which is used as the backbone Feature $\widehat{Y}$ in Equ.(11). Besides, another temporal network, e.g. a Long Short Term Memory Network, exploits the information of variations in the input sequence. As a result, those dynamic information in the video sequence is used as the compensation item of ΔY [87].

4.3.3. Attention is all you need

The Attention mechanism is one of the hottest topic in the field [88]. The relationships between feature maps, pixels in an image, or words in a sentence,

are obtained by the Scaled Dot-Product Attention Mechanism, which is realized by the calculation of three matrices, i.e. the Query, Keys, and Values [6]. Then, the Encoder-Decoder network is compensated with an attention feedback, so comes the *Transformer*. The Attention mechanism constructs the compensation item of ΔY by a concentration on the key areas in the image or voice, which has been widely used in the field, e.g. the Image Caption task [89].

4.3.4. Fine-turn in GPT

Although GPT models have attract a lot of interest [47, 48, 49], Machine Intelligence still has a big gap compared with Human Intelligence [90]. There are still some untruthful or even toxic solutions in the GPT output. In the latest GPT developments, these problems are corrected by a fine-tuning human feedback, e.g. Reinforcement Learning from Human Feedback [91]. The compensation item of ΔY is realized by a fine-turn based on reinforcement learning. Firstly, the recognition model is trained with a supervised learning framework, i.e. the backbone of $\widehat{Y}$ in Equ.(11). Then, the reward model is added to compare the model output with human label. Finally, the reinforcement learning policy against reward model is optimized as fine-turning, so that the open loop recognition system is closed with a human-in-the-loop.

4.4. Discussion on SOTA

Table 2 shows some improvements with a compensation, marked with $+\mathbf{C}$, on several state-of-the-art object detection models. Besides, a comparison of Neural Image Caption (NIC) with or without Attention mechanism is shown in Table 3. Those open problems summarized in Section 3 still merit a further study. The realization of a disturbance rejection control in pattern recognition still lacks a theoretical development on the disturbance observation and compensation. Besides, many successful control strategies have not yet been applied into the recognition tasks. The compensation on features is still an interesting topic. Moreover, the interpretability of all these methods still need more theoretical developments and practical verification.

Table 2: Performances of some state-of-the-art detectors with or without compensation on MS COCO test-dev split [23, 59, 67]. APs: Average Precision related metrics in an object detection task.

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
One-stage detectors							
RetinaNet	ResNet-50	35.7	55.5	38.2	20.4	39.7	46.7
FCOS	ResNet-101	41.0	60.7	44.1	24.0	44.1	51.0
Two-stage detectors							
Faster R-CNN	ResNet-101	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN +FPN	ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2
Mask R-CNN	ResNet-101	38.2	60.3	41.7	20.1	41.1	50.2
Detectors with compensation							
RetinaNet+C	ResNet-50	38.0(**+2.3**)	57.2(**+1.7**)	40.5(**+2.3**)	21.9(**+1.5**)	41.5(**+1.8**)	49.6(**+2.9**)
FCOS+C	ResNet-101	**45.1**(**+4.1**)	**63.9**(**+3.2**)	**49.0**(**+4.9**)	**26.5**(**+2.5**)	**49.1**(**+5.0**)	**57.9**(**+6.9**)
Faster RCNN+C	ResNet-101	40.3(**+4.1**)	62.5(**+3.4**)	43.9(**+4.9**)	22.7(**+4.5**)	43.3(**+4.3**)	51.9(**+3.7**)
Mask RCNN+C	ResNet-101	41.1(**+2.9**)	62.9(**+2.6**)	44.9(**+3.2**)	23.0(**+2.9**)	44.0(**+2.9**)	52.6(**+2.4**)

Table 3: Performances of NIC models with or without attention [89, 92]. B@1, $\cdots$, B@4: Bilingual Evaluation Understudy in n-gram precision ($n = 1, \cdots, 4$) [93]. R: Recall-Oriented Understudy for Gisting Evaluation [94]

NIC Method	B@1	B@2	B@3	B@4	R
MSCOCO					
Without Attention	0.625	0.450	0.321	0.229	0.554
Soft Attention	0.706	0.493	0.344	0.243	0.637
Hard Attention	0.717	0.502	0.358	0.250	0.655
Adaptive Attention	**0.741**	**0.580**	**0.439**	**0.332**	**0.705**
Flickr30K					
Without Attention	0.663	0.423	0.277	0.183	0.554
Soft Attention	0.667	0.433	0.288	0.191	0.607
Hard Attention	0.667	0.439	0.296	0.198	0.606
Adaptive Attention	**0.741**	**0.580**	**0.439**	**0.284**	**0.657**

5. Conclusion

The topic of disturbance rejection in pattern recognition is studied from a system control point of view in this paper. A review on the current technologies is summarized at first. Groups of studies point to a robustness problem in pattern recognition. That is, the internal and external disturbances, which affect the system stability, are still largely understudied. There is still limited research on the disturbance observation and compensation in the field, not only in the recognition model but also in the working system. Therefore, the pattern recognition task is further studied within a compensation framework in this paper. Currently, it is still difficult to unify the recognition process into a series of explicit function similar to the differential equation in a dynamic system. As a result, the realization of disturbance observation and corresponding compensation in pattern recognition is quite different from the dynamic system cases in the industry. The open problems are put forward. Then the potential solutions are summarized based on some current studies in both of the control science field and the pattern recognition field. A new systematic research point of view is developed with the comprehensive review in this paper. These latest developments on disturbance observation and compensation can open up further researches in the field.

Acknowledgments

This work was partially financially supported by Shanghai Cross-disciplinary Research Foundation under grant JYJC202214 and the National Natural Science Foundation of China under grants 61533012, 52041502, 62001286 and 91748120.

References

[1] G. R. Alexandre, J. M. Soares, G. A. Pereira Thé, Systematic Review of 3D Facial Expression Recognition Methods, Pattern Recognition 100 (2020) 107108.

[2] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G. Yang, XAI—Explainable Artificial Intelligence, Science Robotics 4 (37) (2019) 1–2.

[3] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[4] C. J. Liu, H. Wechsler, Gabor Feature Based Classification Using The Enhanced Fisher Linear Discriminant Model for Face Recognition, IEEE Transactions on Image Processing 11 (2002) 467–476.

[5] R. Girshick, Fast R-CNN, in: International Conference on Computer Vision, 2015, pp. 1440–1448.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All You Need, in: Advances in Neural Information Processing Systems, 2017, pp. 6000–6010.

[7] M. I. Jordan, T. M. Mitchell, Machine Learning: Trends, Perspectives, and Prospects, Science 349 (6245SI) (2015) 255–260.

[8] Y. Wang, D. Gong, Z. Zhou, X. Ji, H. Wang, Z. Li, W. Liu, T. Zhang, Orthogonal Deep Features Decomposition for Age-invariant Face Recognition, in: European Conference on Computer Vision, 2018, pp. 764–779.

[9] M. Xue, X. Duan, W. Liu, Eliminati Other-Race Effect for Multi-Ethnic Facial Expression Recognition, Mathematical Foundations of Computing 2 (1) (2019) 43–53.

[10] S. Zafeiriou, C. Zhang, Z. Zhang, A Survey on Face Detection in The Wild: Past, Present and Future, Computer Vision and Image Understanding 138 (2015) 1–24.

[11] T. K. Dash, S. Mishra, G. Panda, S. C. Satapathy, Detection of COVID-19 from Speech Signal using Bio-inspired based Cepstral Features, Pattern Recognition 117 (2021) 107999.

[12] Y. Ge, B. Li, Y. Zhao, W. Yan, HH-Net: Image Driven Microscope Fast Auto-Focus with Deep Neural Network, in: International Conference on Biomedical Engineering and Technology, 2019, pp. 180–185.

[13] P. Delgado-Santos, R. Tolosana, R. Guest, F. Deravi, R. Vera-Rodriguez, Exploring Transformers for Behavioural Biometrics: A Case Study in Gait Recognition, Pattern Recognition 143 (2023) 109798.

[14] L. Claussmann, M. Revilloud, D. Gruyer, S. Glaser, A Review of Motion Planning for Highway Autonomous Driving, IEEE Transactions on Intelligent Transportation Systems 21 (5) (2020) 1826–1848.

[15] E. Huellermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine Learning 110 (3) (2021) 457–506.

[16] C. Hong, J. Yu, J. Zhang, X. Jin, K. Lee, Multimodal Face-Pose Estimation with Multitask Manifold Deep Learning, IEEE Transactions on Industrial Informatics 15 (7) (2019) 3952–3961.

[17] S. Z. Li, R. Chu, S. Liao, L. Zhang, Illumination Invariant Face Recognition Using Near-infrared Images, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 627–639.

[18] X. Hu, E. Guan, W. Yan, Y. Zhao, A Real-Time Abnormal Data Detecting Strategy for Length Sensors Measurement, in: IEEE International Conference on Real-time Computing and Robotics, 2018, pp. 508–513.

[19] J. Shi, C. Qi, From Local Geometry to Global Structure: Learning Latent Subspace for Low-resolution Face Image Recognition, IEEE Signal Processing Letters 22 (5) (2015) 554–558.

[20] A. Kamil, H. Al Ali, D. Dean, B. Senadji, G. R. Naik, Enhanced Forensic Speaker Verification Using A Combination of DWT and MFCC Feature Warping in the Presence of Noise and Reverberation Conditions, IEEE Access 5 (99) (2017) 15400–15413.

[21] Z. Zhang, J. Geiger, J. Pohjalainen, A. E. D. Mousa, W. Jin, B. Schuller, Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments, ACM Transactions on Intelligent Systems and Technology 9 (5) (2018) 1–28.

[22] H. Caesar, J. Uijlings, V. Ferrari, COCO-Stuff: Thing and Stuff Classes in Context, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1209–1218.

[23] X. Hu, W. Xu, Y. Gan, J. Su, J. Zhang, Towards Disturbance Rejection in Feature Pyramid Network, IEEE Transactions on Artificial Intelligence 4 (4) (2023) 946–958.

[24] H. S. Tsien, Engineering Cybernetics, McGraw-Hill, New York, 1954.

[25] N. Minorsky, Directional Stability of Automatically Steered Bodies, Journal of the American Society of Naval Engineers 34 (2) (1922) 280–309.

[26] J. Han, From PID to Active Disturbance Rejection Control, IEEE Transactions on Industrial Electronics 56 (3) (2009) 900–906.

[27] K. Ohishi, M. Nakao, K. Ohnishi, K. Miyachi, Microprocessor-Controlled DC Motor for Load-Insensitive Position Servo System, IEEE Transactions on Industrial Electronics 34 (1) (1987) 44–49.

[28] H. S. Ramirez, On The Dynamical Sliding Mode Control of Nonlinear Systems, International Journal of Control 57 (5) (1993) 1039–1061.

[29] P. Lu, Y. Li, L. Jin, S. Han, Blind Image Quality Assessment Based on Wavelet Power Spectrum in Perceptual Domain, Transactions of Tianjin University 22 (6) (2016) 596–602.

[30] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training, in: IEEE International Conference on Computer Vision, 2017, pp. 2764–2773.

[31] L. Wang, J. Su, K. Zhang, Cross-Database Facial Expression Recognition with Domain Alignment and Compact Feature Learning, in: International Symposium on Neural Networks, 2019, pp. 341–350.

[32] R. Lienhart, J. Maydt, An Extended Set of HAAR-like Features for Rapid Object Detection, in: International Conference on Image Processing, 2002, pp. 900–903.

[33] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: Common Objects in Context, in: European Conference on Computer Vision, 2014, pp. 740–755.

[34] L. Yann, Y. Bengio, G. Hinton, Deep Learning, Nature 521 (7553) (2015) 436–444.

[35] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed Optimization and Statistical Learning via The Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning 3 (2010) 1–122.

[36] J. Goodman, Classes for Fast Maximum Entropy Training, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, pp. 561–564.

[37] K. Wang, W. Zhu, P. Zhong, Robust Support Vector Regression with Generalized Loss Function and Applications, Nature Processing Letters 41 (1) (2015) 89–106.

[38] Z. Feng, L. Jin, D. Tao, S. Huang, DLANet: A Manifold-learning-based Discriminative Feature Learning Network for Scene Classification, Neurocomputing 157 (2015) 11–21.

[39] A. Kendall, Y. Gal, What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, in: Advances in Neural Information Processing Systems, 2017, pp. 5580–5590.

[40] P. D. Sai Manoj, S. Amberkar, S. Rafatirad, H. Homayoun, Efficient Utilization of Adversarial Training towards Robust Machine Learners and Its Analysis, in: IEEE ACM International Conference on Computer-Aided Design, 2018, pp. 1–6.

[41] B. Pes, Learning From High-Dimensional Biomedical Datasets: The Issue of Class Imbalance, IEEE Access 8 (2020) 13527–13540.

[42] Y. Bengio, A. Courville, P. Vincent, Representation Learning: A Review and New Perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 1798–1828.

[43] J. Yim, D. Joo, J. Bae, J. Kim, A Gift from Knowledge Distillation: Fast Optimization, Network Minimization and Transfer Learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7130–7138.

[44] J. Su, D. V. Vargas, K. Sakurai, One Pixel Attack for Fooling Deep Neural Networks, IEEE Transactions on Evolutionary Computation 23 (5) (2019) 828–841.

[45] Y. Li, Sorting of Coal and Coal Waste with Transferred Deep Kernel Learning, International Journal of Systems, Control and Communications 14 (3) (2023) 274–285.

[46] Z. Li, D. Gong, X. Li, D. Tao, Learning Compact Feature Descriptor and Adaptive Matching Framework for Face Recognition, IEEE Transactions on Image Processing 24 (9) (2015) 2736–2745.

[47] X. Zheng, C. Zhang, P. C. Woodland, Adapting GPT, GPT-2 and BERT Language Models for Speech Recognition, in: IEEE Automatic Speech Recognition and Understanding Workshop, 2021, pp. 162–168.

[48] T. B. Brown, et al., Language Models Are Few-Shot Learners, in: Advances in Neural Information Processing Systems, 2020, pp. 1877–1901.

[49] L. Ouyang, et al., Training Language Models to Follow Instructions with Human Feedback, in: Advances in Neural Information Processing Systems, 2022, pp. 27730–27744.

[50] R. Li, et al., An Effective Data Augmentation Strategy for CNN-Based Pest Localization and Recognition in the Field, IEEE Access 7 (2019) 160274–160283.

[51] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random Erasing Data Augmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020.

[52] C. Ding, D. Tao, Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (4) (2018) 1002–1014.

[53] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[54] S. Huang, C. T. Lin, S. Chen, Y. Wu, P. Hsu, S. Lai, AugGAN: Cross Domain Adaptation with GAN-Based Data Augmentation, in: European Conference on Computer Vision, 2018, pp. 731–744.

[55] Y. Huang, J. Su, Output Feedback Stabilization of Uncertain Nonholonomic Systems with External Disturbances via Active Disturbance Rejection Control, ISA Transactions 104 (2020) 245 – 254.

[56] C. Luo, W. Xu, C. Zhu, Robust Gait Recognition Based on Partitioning and Canonical Correlation Analysis, in: IEEE International Conference on Imaging Systems and Techniques, 2015, pp. 269–273.

[57] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[58] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-scale Image Recognition, in: International Conference on Learning Representation, 2015, pp. 1–14.

[59] W. Xu, Y. Gan, J. Su, Bidirectional Matrix Feature Pyramid Network for Object Detection, in: International Conference on Pattern Recognition, 2021, pp. 8000–8007.

[60] H. Yan, Transfer Subspace Learning for Cross-dataset Facial Expression Recognition, Neurocomputing 208 (SI) (2016) 165–173.

[61] G. Xiang, J. Su, A Heuristic Algorithm for Robustly Stable Generalized Disturbance Observer Synthesis with Closed Loop Consideration, ISA Transactions 90 (2019) 147–156.

[62] B. B. Alagoz, F. N. Deniz, C. Keles, N. Tan, Disturbance Rejection Performance Analyses of Closed Loop Control Systems by Reference to Disturbance Ratio, ISA Transactions 55 (2015) 63–71.

[63] D. N. Ngoc, N. Thanh, S. Nahavandi, System Design Perspective for Human-Level Agents Using Deep Reinforcement Learning: A Survey, IEEE Access 5 (2017) 27091–27102.

[64] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A Discriminative Feature Learning Approach for Deep Face Recognition, in: European Conference on Computer Vision, 2016.

[65] R. Khalil, E. Jones, M. Babar, T. Jan, M. Zafar, T. Alhussain, Speech Emotion Recognition Using Deep Learning Techniques: A Review, IEEE Access 7 (2019) 117327–117345.

[66] M. El Ayadi, M. S. Kamel, F. Karray, Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases, Pattern Recognition 44 (3) (2011) 572–587.

[67] Y. Gan, W. Xu, J. Su, SFPN: Semantic Feature Pyramid Network for Object Detection, in: International Conference on Pattern Recognition, 2021.

[68] Y. Qian, W. Deng, J. Hu, Unsupervised Face Normalization with Extreme Pose and Expression in the Wild, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9843–9850.

[69] Q. Lei, A. Jalal, I. S. Dhillon, A. G. Dimakis, Inverting Deep Generative Models, One Layer at A Time, in: Advances in Neural Information Processing Systems, 2019, pp. 13933–13942.

[70] W. Chen, J. Yang, L. Guo, S. Li, Disturbance-Observer-Based Control and Related Methods–An Overview, IEEE Transactions on Industrial Electronics 63 (2) (2016) 1083–1095.

[71] K. Simonyan, A. Zisserman, Two-Stream Convolutional Networks for Action Recognition in Videos, in: Advances in Neural Information Processing Systems, 2014, pp. 568–576.

[72] W. Xue, Y. Huang, On Frequency-domain Analysis of ADRC for Uncertain System, in: American Control Conference, 2013, pp. 6637–6642.

[73] L. Wang, J. Su, Disturbance Rejection Control for Non-minimum Phase Systems with Optimal Disturbance Observer, ISA Transactions 57 (2015) 1–9.

[74] E. Sariyildiz, K. Ohnishi, A Guide to Design Disturbance Observer, Journal of Dynamic Systems Measurement and Control 136 (2) (2014) 1–10.

[75] J. Zhang, L. Guo, Theory and Design of PID Controller for Nonlinear Uncertain Systems, IEEE Control Systems Letters 3 (3) (2019) 643–648.

[76] R. Barbosa, J. Machado, I. Ferreira, Tuning of PID Controllers Based on Bode's Ideal Transfer Function, Nonlinear Dynamics 38 (1-4) (2004) 305–321.

[77] Y. Duan, J. Lu, J. Feng, J. Zhou, Learning Rotation-Invariant Local Binary Descriptor, IEEE Transactions on Image Processing 26 (8) (2017) 3636–3651.

[78] S. Jia, G. Guo, Z. Xu, A Survey on 3D Mask Presentation Attack Detection and Countermeasures, Pattern Recognition 98 (2020) 1–13.

[79] X. Zhang, S. Hu, H. Zhang, X. Hu, Full Occlusion Handling for Pedestrian Tracking via Hybrid System, Turkish Journal of Electrical Engineering and Computer Sciences 25 (2) (2017) 820–831.

[80] R. Bickel, M. Tomizuka, Passivity-based versus Disturbance Observer Based Robot Control: Equivalence and Stability, Journal of Dynamic Systems Measurement and Control 121 (1) (1999) 41–47.

[81] H. Shim, N. H. Jo, An Almost Necessary and Sufficient Condition for Robust Stability of Closed-loop Systems with Disturbance Observer, Automatica 45 (1) (2008) 296–299.

[82] B. Guo, Z. Zhao, Weak Convergence of Nonlinear High-Gain Tracking Differentiator, IEEE Transactions on Automatic Control 58 (4) (2013) 1074–1080.

[83] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent Advances in Convolutional Neural Networks, Pattern Recognition 77 (2018) 354–377.

[84] Y. Saito, S. Takamichi, H. Saruwatari, Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks, IEEE-ACM Transactions on Audio Speech and Language Processing 26 (1) (2018) 84–96.

[85] K. Oksuz, B. C. Cam, S. Kalkan, E. Akbas, Imbalance Problems in Object Detection: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (10) (2021) 3388–3415.

[86] G. Moody, R. Mark, The Impact of The MIT-BIH Arrhythmia Database, IEEE Engineering in Medicine and Biology Magazine 20 (3) (2001) 45–50.

[87] X. Hou, S. Qin, J. Su, Visual Detection of Parkinson's Disease via Facial Features Recognition, in: Proceedings of Chinese Intelligent Automation Conference, 2022, pp. 249–257.

[88] R. Karthik, R. Menaka, H. M, D. Won, Contour-enhanced Attention CNN for CT-based COVID-19 Segmentation, Pattern Recognition 125 (2022) 108538.

[89] D. Qiu, J. Su, Chinese Image Caption Based on Transformer, in: Chinese Control Conference, 2022, pp. 748–752.

[90] J. Su, Y. Chen, Z. Ma, Y. Huang, G. Xiang, R. Chen, From AlphaGo to BetaGo - Quantitative Realization of Qualitative Artificial Intelligence Based on Task Realizability Analysis, Control Theory and Applications 33 (2016) 1572–1583.

[91] N. Stiennon, et al., Learning to Summarize from Human Feedback, in: Advances in Neural Information Processing Systems, 2020, pp. 3008–3021.

[92] K. Xu, J. L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. S. Zemel, Y. Bengio, Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, in: International Conference on Machine Learning, 2015, pp. 2048–2057.

[93] K. Papineni, S. Roukos, T. Ward, W. Zhu, BLEU: A Method for Automatic Evaluation of Machine Translation, in: Annual Meeting on Association for Computational Linguistics, 2002, pp. 311–318.

[94] C. Lin, ROUGE: A Package for Automatic Evaluation of Summaries, in: Annual Meeting on Association for Computational Linguistics, 2004, pp. 74–81.