封面页

参赛队员姓名：花弘扬
中学：北京二中
省份：北京
国家/地区: 中国
指导教师姓名: 高凯
指导老师单位: 北京二中
论文题目：**Re-Designing Road Signs for Autonomous Driving?** **More Recognizable Signs Designing via Simulated Annealing**

Re-Designing Road Signs for Autonomous Driving?

More Recognizable Signs Designing via Simulated Annealing

Hongyang Hua[1]
Kai Gao (Mentor)[1]

1 Beijing No.2 Middle School
Beijing, China

Abstract

Road sign recognition under high-speed driving scenarios faces numerous challenges, such as motion blur and low resolution. Traditional approaches tend to focus on improving the capability of recognition networks, which may result increased inference time. From a different perspective, we propose a novel mathematical problem: how to design a road sign system that is more easily recognizable by computers in complex environments.

To address this problem, we simplify the road sign design task into a combinatorial problem. Specifically, we investigate the possibility of selecting M road signs from a given candidate set to minimize the corresponding training-testing error rate. To achieve this, we combine machine learning and simulated annealing to provide an optimal road sign selection solution. Our method considers M sets of candidate signs, K choices for each set, encompasses the relationships among KM distinct road signs, and ultimately identifies M highly distinguishable signs. Experimental results demonstrate that our proposed road sign design solution achieves excellent recognition accuracy in various blurry scenarios.

The road sign design solution presented in this paper can provide valuable reference suggestions for the government in designing new national standard road signs. Additionally, in specific enclosed environments such as ports and mines, this solution can enhance the reliability of autonomous driving technology. Therefore, in the era of widespread adoption of autonomous driving systems, our solution can be widely applied to mitigate traffic accidents caused by road sign recognition errors. Furthermore, the new problem and corresponding algorithm introduced in this paper can find applications in various other scenarios, such as character design in animation.

摘　要

在高速驾驶的场景中，路牌识别面临着各种挑战，如运动模糊和低分辨率等。传统的研究者往往致力于提升识别网络的能力，然而会随之带来推断时间下降等问题。本文从另一个角度出发，提出了一个新的数学问题:即如何设计适应于复杂环境下，更容易被计算机识别的路牌体系。

进一步，本文将路牌设计问题简化为一个组合问题，即能否从一定的候选集中，选择出 M 个路牌，使得其对应的训练-测试错误率能够达到最小。并且围绕这样一个问题，结合了机器学习和模拟退火，给出了一套最优的路牌选择方案.该方法考虑了 M 组候选路牌，即 3M 种路牌之间的相互关系，最终得到 M 个最适合的路牌，相互之间尽可能不混淆。实验结果显示，本文提出的路牌设计方案在各种模糊场景下都具有优良的识别准确率。

本文提出的路牌设计方案，可以为政府设计新的国标路牌提供参考建议。另外在特定的封闭环境中，如港口和矿场等，该方案也可以提高自动驾驶技术的可靠性。因此，在自动驾驶系统普及的时代，该方案可以被广泛应用，降低因路牌识别错误而导致的交通事故发生率。同时本文提出的新问题和对应的算法，也可以在更多的场景（如动漫人物设计等）中使用。

1 Introduction

According to an analysis by the US Department of Transportation in 2018, 94% of traffic accidents are caused by human drivers. In order to reduce the occurrence of traffic accidents, autonomous driving systems and advanced driver assistance systems have been rapidly developed. Many car manufacturers, such as Honda, Toyota, Mercedes-Benz, BMW, and Audi, have emphasized the use of technologies related to autonomous driving and driver assistance. These technologies utilize cameras to recognize speed limit signs and provide prompts or assistance to the drivers [1-3].

As visual information is a fundamental component of existing transportation systems, the key functionalities of all current autonomous driving systems rely on cameras to provide visual information for road condition recognition, complemented by LIDAR and radar to prevent collisions. In terms of road condition recognition, the recognition of road signs is one of the important problems that need to be addressed.

In machine learning-related courses, we learn that machine learning requires learning the distribution of samples through a dataset and building a model for recognition. Researchers in the field of artificial intelligence often focus on improving the algorithm capacity to enhance the accuracy of recognizing images captured by cameras using existing given road signs. For instance, Richard Xue's work on CT classification, which received the silver award in the 2021 Yau Awards, adopted a multi-angle and multi-task training approach to improve performance. However, there are still many challenges in this approach: 1) Many road signs are inherently confusing in design, with numerous signs that have similar appearances but drastically different meanings, leading to confusion; 2) Many road signs have complex patterns, making critical information difficult to discern from a distance; 3) Existing methods to address low resolution or blurring in recognition mostly rely on deep learning networks, which have limited real-time capabilities and compromise safety to some extent; 4) Existing deep learning-based road sign recognition systems are mostly trained on existing data and may not possess strong transferability when encountering new or specialized road signs, requiring retraining to achieve good recognition models, which is costly.

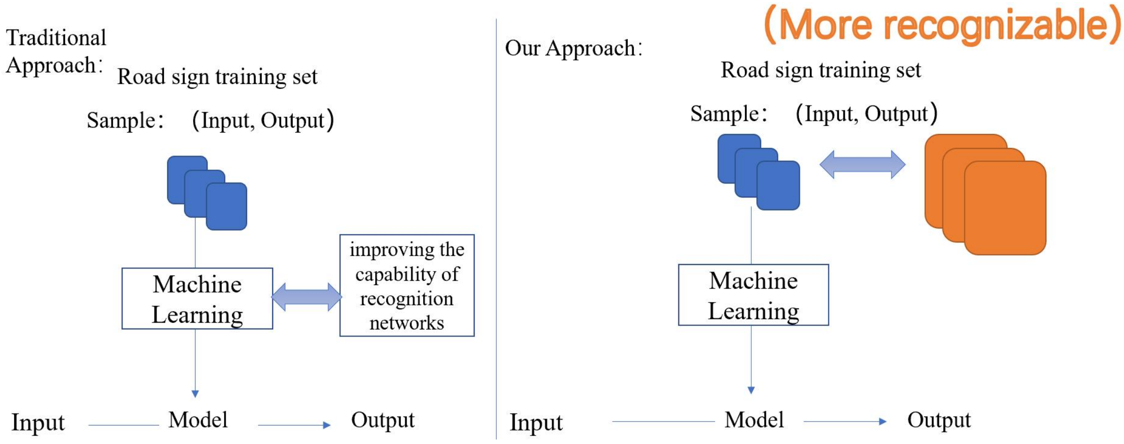


Figure 1 Instead of enhancing the classifiers of machine learning, this project discusses how to obtain more recognizable classification objectives under a specific application (such as autonomous driving).

Therefore, I discussed with my advisor the possibility of improving the overall accuracy of the system by changing the design of road signs used in road sign recognition. Although my advisor mentioned that this strategy may not be applicable to self-driving companies, it could be considered by the government to revise the road signs and replace them with a more machine-readable set of signs, thereby enhancing the accuracy of autonomous and assistive driving systems for the benefit of society as a whole. Furthermore, during China Adolescents Science & Technology Innovation Contest, experts also highlighted the potential of our approach to design road markers for internal environments such as mines and ports. It is worth noting that this project presents a novel algorithmic problem and provides an initial solution, which involves the design and selection of training labels that could be applied to domains beyond autonomous driving. In section 5.6, we demonstrate this approach using the example of designing cartoon characters.

Instead of focusing on improving the accuracy of road sign recognition, this study aims to address the issue from the source of the data itself. By enhancing the design feasibility of road signs and devising a more discernible road sign solution, we aim to enhance the recognition capability of autonomous driving systems. Through a series of research endeavors, we aspire to develop a systematic methodology for road sign design, allowing us to create a set of road signs based on this approach. By doing so, we can significantly improve the recognition accuracy of road signs for both autonomous and assistive driving systems, consequently elevating the overall safety of intelligent driving.

1.1 Challenges in Road Sign Design

In the spring of 2023[1], we explored the use of generative models like Stable Diffusion for road sign generation. We constructed a dataset of 30 sentences related to road signs and fed them into Stable Diffusion, expecting it to generate new road sign designs. The results, as shown in Figure 2, were as follows:

[1] In the spring of 2023, my teacher from YINGCAIJIHUA (英才计划), a plan jointly organized by the China Association for Science and Technology and the Ministry of Education for cultivating talents for scientific and technological innovation among middle school students, introduced us to the cutting-edge technology of generative models. We also came across videos on platforms like Bilibili, where users showcased images generated using Stable Diffusion. Intrigued, we as high school students decided to explore the use of Stable Diffusion for road sign generation as a class project. This appendix provides background on the educational context of the project.

a road sign indicate that parking car is restricted here	A road sign indicating no bicycles allowed ahead	A road sign showing no motorcycle	A road sign showing there's a circular road

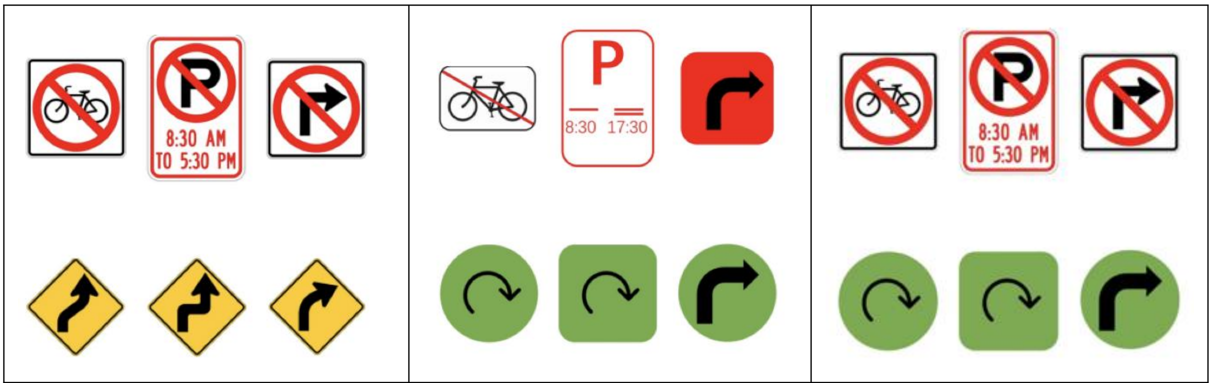
Figure 2 Signs generated using stable diffusion

1. Stable Diffusion occasionally produced interesting road sign designs, but more often it failed to capture the meaning of our instruction.
2. Even when generating road signs, there was no guarantee that the algorithmic output would be easily distinguishable from existing road signs.
3. Modifying the Stable Diffusion algorithm proved to be excessively complex for this project.

Clearly, attempting to make machines directly generate easily recognizable combinations of road signs would be too intricate and challenging. While generative models show promise for creative applications, road sign generation requires further research into controllability and interpretability. We intend to explore techniques like class-conditional guidance to improve the coherence of generated signs in future work.

1.2 Transforming Generation Problem into Selection Problem

In the field of information competitions, various combinatorial optimization problems arise frequently. For example, trucks can choose different nodes to deliver goods, or different countries can dispatch different envoys for diplomatic purposes. The ultimate goal is to find the most optimal solution with maximum overall value. Given this background, can we also consider transforming road sign generation problem into an optimization problem?



Original road signs	Self-designed road signs	Road signs selected by algorithm

Figure 3 Example of road sign selection

As shown in Figure 3, let's assume that we have designed a set of candidate road sign solution in advance. Let M represent the number of road sign types, and for each type, we have designed K candidate road signs, where K represents the number of choices for each road sign type. Therefore, the total number of candidate road sign designs is KM, resulting in K^M possible combinations. Consequently, we can transform the problem of generating easily recognizable road signs into the task of selecting the combination with the highest classification accuracy among the K^M designed road sign combinations.

Hence, road sign design can be regarded as a combination of manual design and optimal selection. However, finding the combination with the highest accuracy among the given combinations is a complex task, requiring an algorithm capable of efficiently searching for the optimal combination. In this study, we employed the Simulated Annealing Algorithm (SAA) to quickly search for the ideal road sign solution in terms of recognition accuracy. We further validated the effectiveness of the searched road sign solution through experiments.

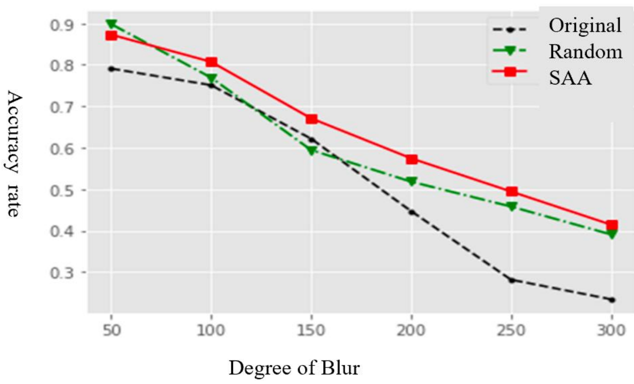


Figure 4-a Recognition accuracy under different blur levels
The horizontal axis is the degree of blur of different traffic signs, and the vertical axis is the recognition accuracy of the model under several solutions. It can be seen that our simulated annealing solution achieves better recognition accuracy under high blur conditions.

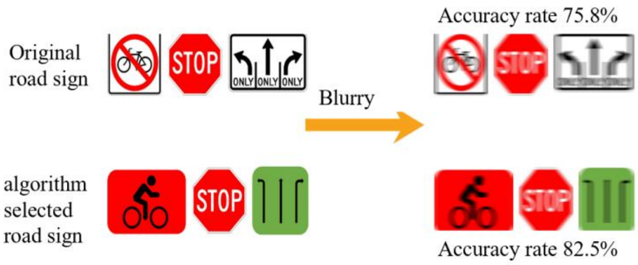


Figure 4-b The contrast of original road sign and algorithm selected road sign under blur degree=100

The classification accuracy rate of algorithm selected road sign outperforms original road sign by 12% in average.

Figure 4-c Head worn display box

Inspired by a presentation method at MIT's workshop, I created a head worn display box, which allows me to experience the blurring of road signs relative to the camera at different positions. Due to the different imaging principles between the human eye and the camera, the images obtained by the camera at high speeds are often more blurry than those captured by the human eye.

The primary contribution of this paper is the significant improvement in classification accuracy of road signs after redesign and selection, which can reduce the probability of traffic accidents caused by misidentification in autonomous driving/assisted driving systems. Specifically, this study presents several innovative aspects:

1. We provide and design a set of road sign solutions with higher classification accuracy, and propose a selection method based on the Simulated Annealing Algorithm (SAA) for choosing the optimal combination of road signs.

2. We break away from the traditional optimization approaches for recognition by introducing a novel idea for improving recognition accuracy, which is to modify the objects being recognized instead of changing the recognition system. Experimental results demonstrate that our algorithm outperforms randomly selected road signs or existing road signs, achieving a 6-20% improvement in recognition accuracy across 55 different road signs under various degrees of blur.

3. To validate the universality of our designed and selected road sign solutions, we conduct cross-model validation experiments, demonstrating significant improvements in recognition results for Model B after implementing the solutions designed by Model A. This validates the generality of our approach and provides a feasible solution to guide the design of optimal strategy selection methods for similar research work.

The remaining sections of this paper are organized as follows: in Section 3, we test the performance degradation of deep learning models, including classification models and detection models, under motion blur. In Section 4, we introduce our road sign design algorithm. In Section 5, we conduct extensive experiments to demonstrate the effectiveness and generalization capability of our proposed method.

2 Related Works

2.1 Visual Detection and Recognition in Autonomous Driving

Visual detection and recognition technologies play a crucial role in autonomous driving systems, especially in the recognition of road signs and vehicles. However, the identification of blurry and low-resolution road signs remains a challenging problem. Previous studies have indicated that damages, low resolution, and motion blur are the main causes of recognition errors in road signs [4-6]. Among them, addressing the issue of low-resolution and motion blur in road sign recognition is of particular importance due to the difficulty of mitigating damages problem through other technical means.

2.2 Recognition of Blurry and Low-Resolution Images

Blurry and low-resolution images are common types of image degradation caused by factors such as distance, focus blur, and motion. There are two main approaches to address the recognition of low-resolution images: one is to restore the image resolution before recognition [7], and the other is to directly train recognition models using blurry images [8]. However, both methods can only partially alleviate the decrease in recognition accuracy caused by low pixel density, and their performance rapidly deteriorates with increasing levels of blur.

2.3 Road Sign Redesign

Some car manufacturers have proposed improving road sign materials to enhance visibility, especially during nighttime. Some research has suggested using electronic signs to communicate information to vehicles and assist in road sign recognition [9]. However, these methods have not been widely applied due to cost considerations and other constraints. Although existing AI-based methods can generate road signs, the generated signs often suffer from poor legibility, as shown in Figure 2.

3 Challenges in Motion Blur Road Sign Recognition

The performance of a basic recognition system can deteriorate when it is affected by factors such as motion blur. In this section, we aim to experimentally verify that even with the use of advanced deep learning methods, the accuracy of recognition can still experience significant declines.

3.1 Qualitative Observations on Real-World Images

We collected experimental data from real-life scenarios using an Apple 11 (12-megapixel) smartphone positioned at the passenger seat. Two speed limit road signs encountered during the journey are shown in Figure 5. With the naked eye, it is almost possible to determine that the road sign on the right indicates a speed limit of 30 km/h (though it could also be mistaken for 20 km/h). However, machines may struggle to accurately identify this sign, resulting in a drop in classification accuracy. On the other hand, the road sign on the left is completely blurred and illegible.

Consider the following scenario: the distance from the blurred road sign to the front of the car in the direction of motion is approximately 20 meters (along the direction of the road), and the lateral distance from the car to the roadside where the sign is located is approximately 10 meters. Assuming a car speed of 50 km/h, the driver's reaction time is only 0.5 seconds, corresponding to a distance of about 7 meters traveled. It is challenging for a car moving at 14 m/s to react within the remaining distance, which greatly increases the risk of accidents. Even the relatively clear road sign on the right in the captured image is not considered clear when observed under a camera. If misidentified as 80 km/h, it could lead to even more hazardous situations in the context of autonomous driving or when the vehicle needs to perform auxiliary analysis.

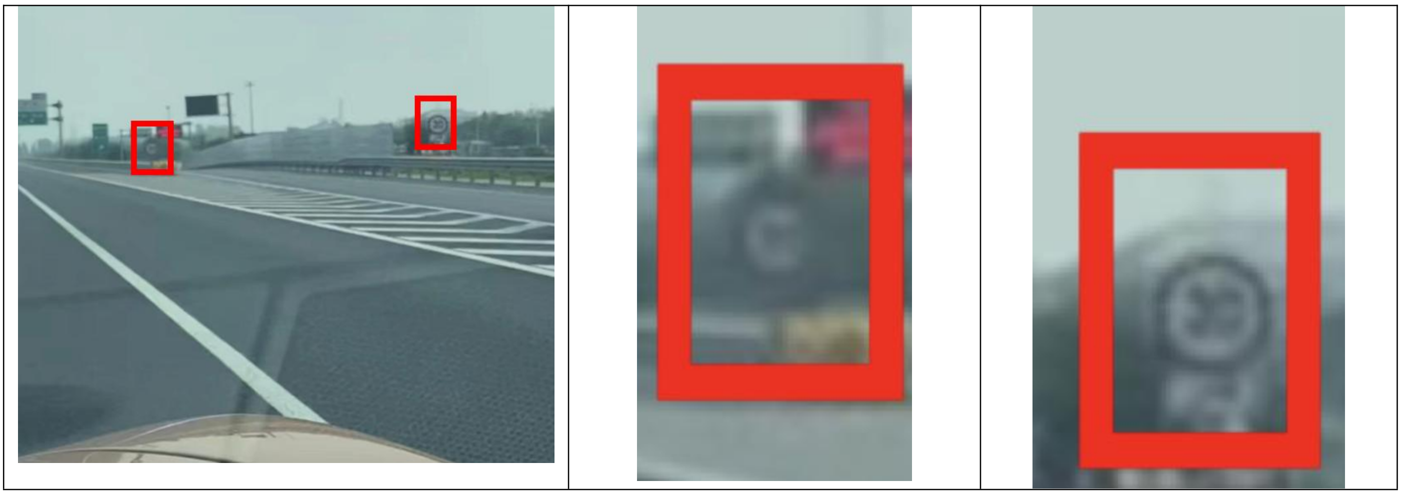


Figure 5 Road sign capture on highway

3.2 Traffic Sign Recognition Baseline

To establish a benchmark for our research and validate the effectiveness of our classification methods, we first need to recognize clear road signs. In this study, we utilized the MobileNet-v2 model to perform preliminary training and validation on a traffic sign dataset consisting of 55 types. The road signs used in this dataset were obtained from the Lisa Traffic Sign Dataset, which features American road signs.

Due to the limited data available, we applied data augmentation techniques to the road sign images. Considering the unique characteristics of road signs, operations such as flipping were not applicable. Instead, augmentation mainly involved adjusting brightness, selecting minor angles, and applying slight affine transformations. 20% of each type of road sign was set aside as a test set. For our research, we employed the pre-trained MobileNet-v2 model, which is a lightweight model designed for mobile devices. Despite its faster training speed compared to ResNet50, it maintains a high level of accuracy. The end-to-end training in our study was based on this model architecture.

The results of the tests showed that the accuracy obtained by directly extracting features was 76.3%, while fine-tuned model increased the accuracy to 87.8%. This finding indicates that fine-tuned model leads to higher accuracy in road sign recognition. The following experiments will also use fine-tuned model as the final standard. Furthermore, as shown in Figure 6, t-SNE can be employed to visualize the feature space. t-SNE is a data visualization tool that reduces high-dimensional data to 2-3 dimensions while preserving inter-feature distances. The t-SNE results demonstrated that road signs trained using our end-to-end approach were more dispersed in the feature space. This means that even when the signs were blurred, distinguishable spatial differences can still be observed among different road signs. Therefore, fine-tuning training indeed improves the accuracy of road sign recognition.

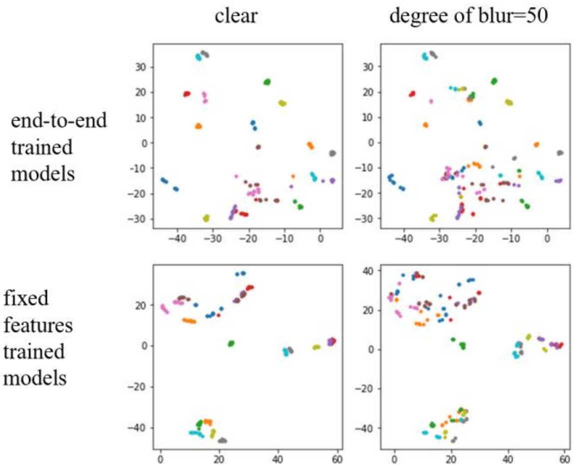


Figure 6 Visualization results of t-SNE using existing model and fine-tuned model under different degrees of blur

The low accuracy of road sign recognition stems from several factors. Firstly, clear road signs undergo significant changes, such as rotation, which hinders accuracy. Additionally, the network only performs fine-tuning, resulting in limited data, especially for road signs with only numerical differences. This paper specifically focuses on using a pre-trained image classification model for recognition instead of specific identification of similar road signs, leading to instances of misidentification. Our experiments utilize the MobileNet-v2 as the training model, in line with the focus on road sign design. All discussions are based on the MobileNet-v2 model. Improving the model structure aligns with the study's design, as a road sign system that facilitates recognition would enhance accuracy. However, during the province-level contest of China Adolescents Science & Technology Innovation Contest, many judges have highlighted the lack of generalization ability in solutions and conclusions derived from training a single model. Therefore, to strengthen our arguments, we will provide experimental results using VGG16, ResNet50, and Swin-Transformer models in the following cross-comparison.

3.3 Blurry Road Sign Recognition

To simulate road sign recognition in complex real-world scenarios, mainly to consider various types of deterioration operations, the operator is denoted as *O* and multiple operations are combined together. In real-world complex scenarios, image blurring is caused by interference. The interference we refer to includes motion blur, focus blur, low ambient light, low-resolution at long distances, and image noise. Due to the continuous movement of the vehicle and the time before opening the shutter, objects that are moving quickly will have traveled a certain distance, and their images will also have moved a distance on the film or image sensor. When multiple images overlap, it leads to a blurry photograph.

For operation *O*, this paper uses four different algorithms to simulate four different types of blurring effects that are related to four practical scenarios that may affect the accuracy of road sign recognition: motion blur caused by acceleration, low-resolution caused by long distances, Gaussian blur caused by defocus, and high noise caused by low ambient light. Figure 7-a demonstrates examples of two common road signs (speed limit sign and U-turn sign) and uses algorithm to simulate the images that a camera may obtain in these situations.

For the four types of deterioration operations, each operation corresponds to a deterioration coefficient. In the case of motion blur, the blur occurs within the range of positive and negative 10 degrees in the horizontal direction. The degree of deterioration is represented by the size of the blur kernel y_m. In the case of low-resolution, when the deterioration degree is y_l, the output image resolution is $40/y_l \times 40/y_l$. For Gaussian blur, the deterioration degree y_g is used as the size of the Gaussian kernel. For adding noise, Gaussian noise is used, and the generated Gaussian noise image is added to the original image with a deterioration degree of y_n.

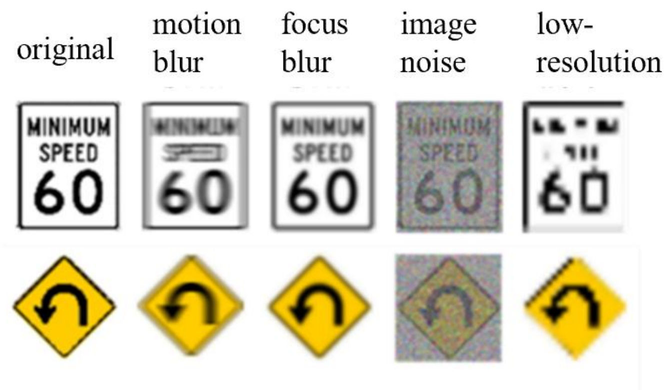


Figure 7-a Schematic diagram of

Figure 7-b Road sign capture

different damage simulations on highway

3.4 Impact of Blurred Road Signs on Recognition

As mentioned earlier, the accuracy of road sign recognition decreases gradually with increasing deterioration levels under various blur conditions. In this study, we conducted experiments to validate this phenomenon. We applied motion blur to the images at different degrees of blur ranging from 50 to 300. The recognition accuracy significantly decreased for both tasks using fixed features and end-to-end trained models. Quantitative results of the accuracy degradation in both classification and detection tasks under different degrees of blur are presented.

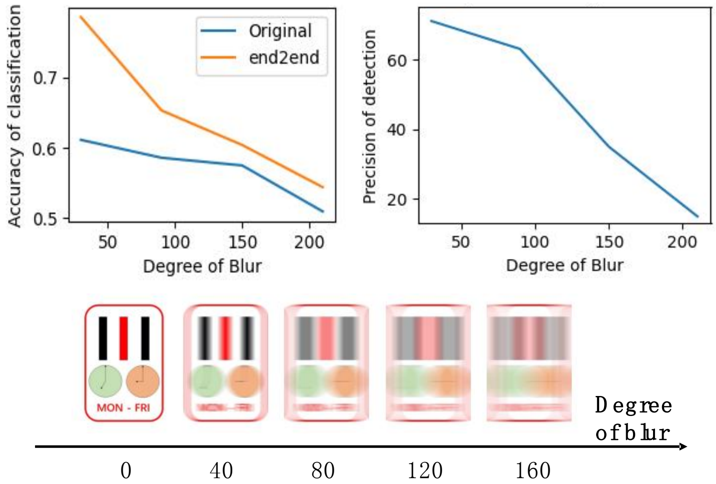


Figure 8 Whether for end-to-end training or fixed features, there will be a severe drop in accuracy under blurry conditions, and the performance of object detection will also suffer a severe decline.

It is evident that severe blur leads to substantial performance deterioration for both classification and detection tasks. Although many existing methods aim to enhance performance by strengthening the network, it is important to note that these methods often increase the inference time and do not conflict with the approach proposed in this paper. Therefore, they can be combined with our method. Our objective is to find a more suitable road sign design solution that can improve accuracy under different blur degrees without modifying the model or training approach. To the best of our knowledge, this is a novel problem that has been rarely investigated in prior research.

4 Road Sign Design and Recognition

4.1 Address the Problem

As mentioned in section 1.2, we have transformed the problem of generating a set of easily recognizable road signs into a problem of finding the ideal solution among K^M combinations of road sign designs. In this case, we consider K=3 and M=55. The ideal solution refers to a set of road sign design solutions that achieve the highest classification accuracy under the presence of blurriness. Specifically, two alternative designs are provided for each road sign in the design process.

Clearly, the most accurate method would involve training and testing the road sign data for each selected design solution. However, even with a simple linear classifier, this would require a significant amount of time. Therefore, we need to find an algorithm that can: 1) quickly evaluate the accuracy of a road sign design solution, and 2) greatly reduce the search space of K^M, thereby minimizing the number of required searches.

Given the large number of potential design solutions for road signs, even in cases with a small number of road sign categories such as M=10, there would be close to sixty thousand possibilities to validate. Conducting a search to find the highest accuracy design solution for each category would be time-consuming. Hence, in this study, we choose to employ a simulated annealing algorithm, a method suitable for searching optimal solutions within a large search space, to approximately obtain the ideal solution.

4.2 Relationship Between Accuracy and Data Distribution in Feature Space

In this project, the features distinguishing road signs include the numbers on speed limit signs and the orientation of turn arrows (as well as color and shape characteristics such as circular, triangular, blue, and yellow). The data points in the feature space correspond to the digitized features of road sign images obtained through a neural network. Classification by the neural network is based on these features. When there is an overlap in the feature point regions, the classifier may not be able to completely separate them, leading to classification errors. The more overlap there is between the feature points of two road signs in the feature space, the higher the error rate of classification.

Figure 9 provides a more intuitive illustration of this phenomenon. Let's assume there are three types of road signs, A, B, and C, each with K=2 design solutions. The smaller the overlap between the designs of each type of road sign, the easier it is for the machine to distinguish them, resulting in higher classification accuracy. However, if there is a larger overlap, it indicates that the designs of these solutions are more similar, and the machine is more likely to make confusions during classification. For example, in Figure 9, solutions A2, C1, and B2 have a higher overlap, resulting in a lower classification accuracy. The machine cannot determine whether it is recognizing A1 or A2 because both designs share certain features. On the other hand, solutions A1, C2, and B2 have no overlap, resulting in higher classification accuracy. The principle is the same for road sign design – it is essential to aim for designs that have smaller overlaps in the feature space, which means higher distinctiveness. The fewer overlaps there are, the more ideal the design solution.

The ideal solution aims for a higher accuracy in road sign classification, which reflects as increased distances between the class centers of any two road sign categories in the feature space, and smaller overlapping areas between two circles. t-SNE visualization can be employed here. After extracting features and visualizing them in the t-SNE feature space, we hope that the new design solutions can maximize the proximity between candidate solutions for each road sign category, as greater distances represent easier distinguishability among road signs.

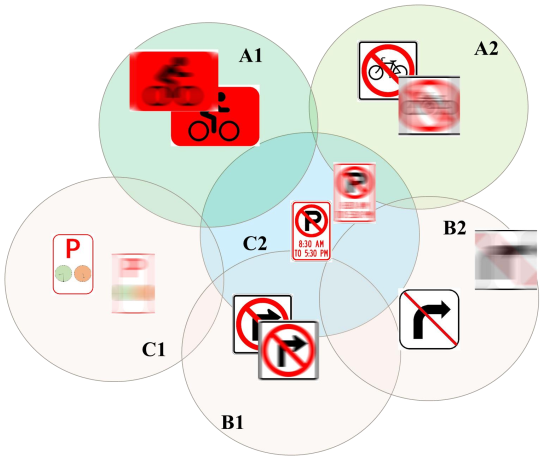


Figure 9 Illustration of feature distribution

The current challenge lies in finding an optimal solution with maximized distances between class centroids using algorithms. It is important to note that the extracted features for any image are of equal length and do not suffer from reduced feature numbers caused by motion blur. Let y represent the extent of image degradation, where a smaller y indicates less image damage. When an object image is subjected to noise or other interference, the extracted features will become blurred. Treating blur as interference, this interference will cause greater variations in feature distributions among different road signs. If y increases, the interference on the image's features also increases, causting the extracted features more different from the clear image features, or the center of the circle, resulting in a lack of distinctiveness in the distribution of data points (feature points) in the feature space.

Simulated annealing is a general probabilistic algorithm used to find the optimal solution for a given proposition within a large search space. The term "simulated annealing" comes from the metallurgy term "annealing". Annealing involves heating a material and then cooling it at a specific rate to increase the volume of grains and reduce defects in the lattice. Originally, the atoms in the material would settle in positions that minimize internal energy. Heating increases the energy, causing atoms to move away from their original positions and randomly explore other positions. During the slow cooling of annealing, atoms have a higher chance of finding positions with lower internal energy.

Simulated annealing simulates this physical process by defining an energy function E and continuously making small modifications to the current state, accompanied by a gradual decrease in temperature T. Through random movements, the algorithm gradually selects status with lower energy E, eventually finding the optimal solution. Taking road signs as an example in this paper, as mentioned earlier, there are 3^{55} different solutions. The simulated annealing algorithm randomly selects an initial solution and continuously modifies the road sign arrangement while attempting to retain the better solutions. It continues to explore based on these solutions. Although it may occasionally produce inferior solutions, with continuous attempts, the solutions obtained through simulated annealing will gradually improve.

Simulated annealing requires evaluating the quality of a set of solutions. If accuracy is used as the evaluation criterion, it would still require testing with a set of data, which is time-consuming. Therefore, an approximation of accuracy can be obtained by measuring the distances between different classes in the feature space as an evaluation criterion.

4.3 Fitting the Relationship between Distance and Accuracy

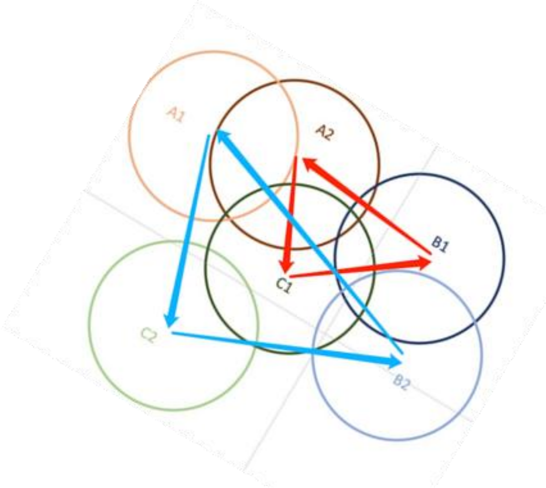


Figure 10 Schematic diagram of distance d

The objective function of simulated annealing is the energy E. In this case, we want E to directly correspond to the classification accuracy. First, we define the distance between images in the feature space (the distance between class centroids, as shown by the arrows in Figure 10) as d. We design a formula to ensure that E and our class centroid distance d have a specific relationship: the smaller the E, the larger the d. This allows us to approximate the relationship between accuracy and the distance d between all classes.

To achieve a smaller error rate P_e, a larger d is preferred, while the energy E decreases during the cooling process of simulated annealing. Therefore, we take E as the error rate P_e. Considering that d is a non-negative value, we let d be the denominator (as the denominator increases and the numerator remains the same, the value will be smaller). The formula for the approximation is as follows, where the error rate P_e equals:

$$P_e = a/d^b . \quad (1)$$

Here, a and b are the parameters to be fitted, and the fitting results are shown in Figure 11.

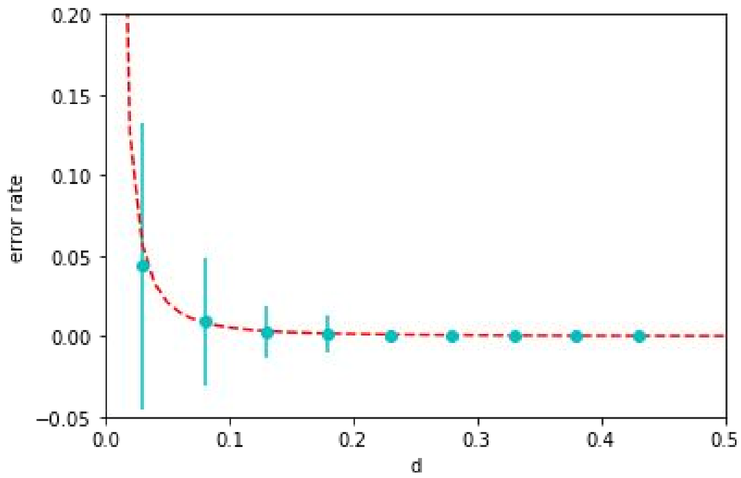


Figure 11 Curve fitted to the relationship between distance and error rate

From the fitting results, we can determine that the energy E satisfies:

$$E = \sum_{ij} a/d_{ij}^b. \quad (2)$$

4.4 Implementation of Simulated Annealing

In the process of using simulated annealing algorithm to search for a reasonable solution, a sign in a combination of signboards with energy E will be randomly changed, and obtain a new combination of signboards with energy E_1. We handle two scenarios as follows: In the first scenario, if E_1 is less than or equal to E, we keep the combination of signboards corresponding to E_1 and continue to randomly change the signboard selection. In the second scenario, if E_1 is greater than E, we select E_1 with a certain probability, which is given by the probability formula P shown below.

It is worth noting that the definition of temperature is particularly important in the simulated annealing algorithm. If the temperature is high, the machine will constantly try different strategies, including both good and bad ones. However, if the temperature is low, the machine will always choose the strategy with the smallest E, leading to getting stuck in a partial minimum. Therefore, we define a temperature T (which decreases over time) to avoid getting stuck in a partial minimum. The commonly used formula in the simulated annealing algorithm is as follows:

$$P = e^{-\Delta E/T} , \quad (3)$$

As time T decreases, P also decreases. In other words, as time goes by, the possibility of choosing E_1 in the second scenario becomes smaller. This means that the chance of conservatively choosing a good strategy increases, which allows us to keep trying different strategies in the early stages and gradually select better strategies later on. This approach not only avoids getting stuck in a local minimum, but also allows us to obtain better strategies than before.

4.5 Replacing distance estimation with fixed feature Learning Accuracy[2]

During model optimization, we adopted simulated annealing with only around 1000 iterations. Given a fixed feature training time of 1 minute, this allowed the algorithm to produce results within a few hours. To further analyze the impact of accuracy estimation, we supplemented experiments with models trained on fixed features. This directly yielded test accuracy, as reported in the Experiments section.

5 Experiments

The experiments section is organized as follows: In section 5.1, we present preliminary results using MobileNet and a basic simulated annealing algorithm, as well as compare the performance difference between using distance estimation and directly using the end-to-end ground truth accuracy. In section 5.2, we address the issue of generalization ability raised by several judges by providing cross-validation results for networks such as VGG and Swin-Transformer, demonstrating the universality of our approach. Section 5.3 is a supplementary experiment to examine the impact of ambiguous directions on recognition accuracy and sign design principles. In section 5.4, as most of the experiments focus on image classification, we include an additional experiment on object detection to demonstrate the improved performance of our sign design in detection tasks. In section 5.5, we attempt to use a sliding rail to capture more realistic data and further validate our approach.

5.1 Experiment

5.1.1 First Impression on Simulated Annealing

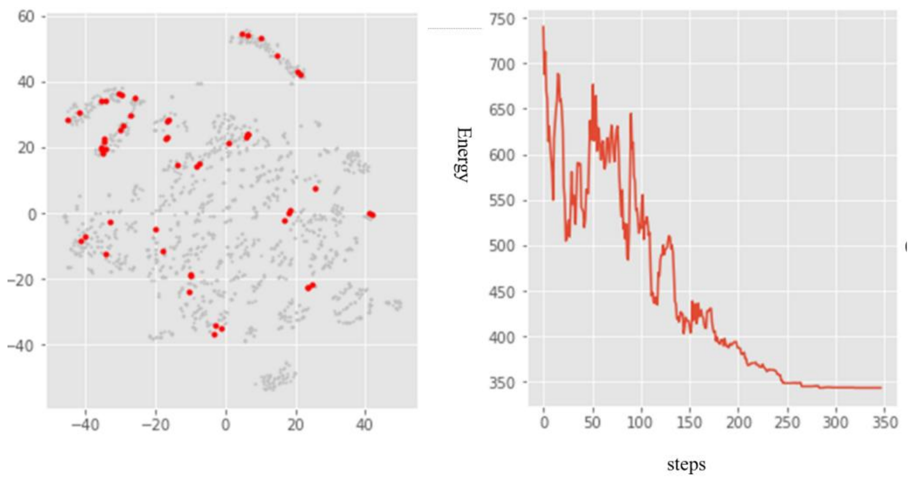


Figure 12 Simulated annealing results

Among the 55 sign types, each sign had two alternative patterns, along with the original sign, forming three options. The center points of each sign were extracted under clear conditions, and simulated annealing was used to find the optimal solution. The distribution of all signs in the feature space was visualized in a two-dimensional plane using the t-SNE algorithm. As shown in the left side of Figure 12(gray dots), some sign patterns were closer and formed clusters in the feature space. The energy E decreasing process during the simulated annealing is shown on the right side of Figure 12. In the early stages of the simulated annealing, when the temperature T was high, the energy E would rise significantly in neighboring cycles. As the temperature T decreased, the probability P also decreased. As time increased, the probability of choosing the most recent step gradually decreased, transitioning from an initial period of random attempts to a stable selection. Setting the temperature allowed for a buffer to find better solutions and extended the time to search for a solution, resulting in a solution that was better than getting stuck in a partial minimum from the beginning. The selected solution, visualized through t-SNE, is shown in the left side of Figure 12(red dots), with each selected sign maintained a certain distance from others.

We further evaluated the accuracy of the sign combinations generated by the simulated annealing algorithm. The results for the original signs, randomly selected sign combinations, and sign combinations selected using simulated annealing under different degrees of blur are shown in Figure 4-a.

[2] At the presentation for the province-level contest of China Adolescents Science & Technology Innovation Contest, one judge pointed out the low number of iterations for our simulated annealing approach. We had only used around 1000 iterations due to time constraints. The judge noted that within a limited timeframe (a few hours), we could still produce results by training with fixed features, since each iteration only required 1 minute. Following this feedback, we supplemented additional experiments by training models on fixed features, in order to directly evaluate test accuracy.

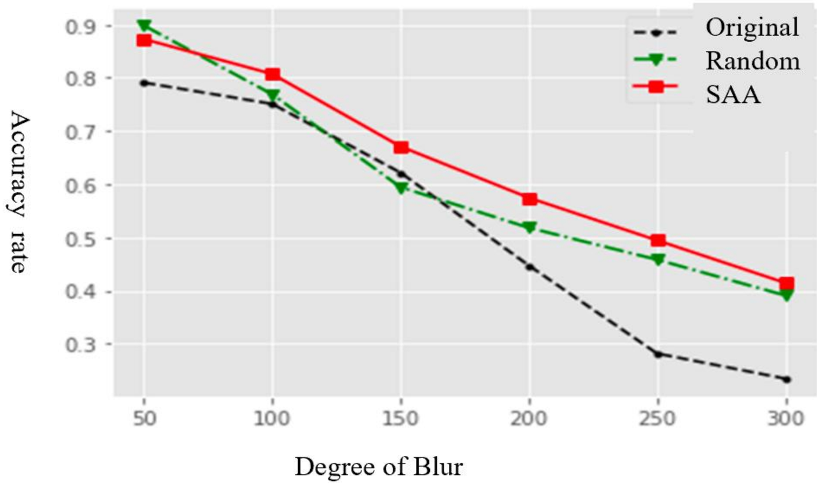


Figure 4-a Recognition accuracy under different degrees of blur

The x-axis represents the degree of blur, while the y-axis represents the classification accuracy. The black line represents the accuracy of the original images, the green line represents the accuracy of randomly selected blurred sign combinations, and the red line represents the accuracy of sign combinations selected using the simulated annealing algorithm. Comparing the red line with the black line, we can observe that after using simulated annealing, the classification accuracy of the sign combinations significantly improved compared to the original solution, with an improvement of nearly 10%. As the degree of blur increased, the accuracy rate increased by nearly 20%. Comparing the green line with the black line, we can see that the accuracy of the randomly selected results is also lower than the results obtained from simulated annealing, indicating that simulated annealing effectively provides a set of sign combinations that are easier to recognize. Through analysis, we can conclude that the new sign design has a certain effect on improving accuracy, and the use of the simulated annealing algorithm is also crucial.

Examples of the two approaches are shown in Figure 4-b. The sign combinations generated by the new approach have higher distinguishability and clearer boundaries under blurred conditions, leading to higher accuracy.

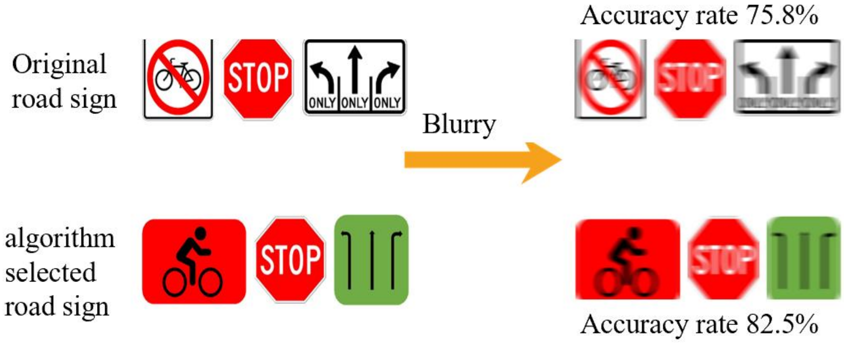


Figure 4-b The contrast of original road sign and algorithm selected road sign under degree of blur =100

The classification accuracy rate of algorithm selected road sign outperforms original road sign by 12% in average.

The classification accuracy of the original signs decreases significantly under conditions of motion blur and long-distance observation, while the sign combinations designed by us reduce the impact of motion blur on sign recognition under the same conditions, thereby improving classification accuracy. As shown in Figure 4-b, the final experimental results demonstrate that our proposed traffic signs have a significant advantage in recognition rate under various blur conditions. In this experiment, Group A consists of a set of existing signs, with degradation conditions including motion blur, focus blur, low environmental illumination, low resolution at long distance, and image noise. After designing two road sign set, Group B is the optimal result selected from the three sets (two designed sets and one original road sign set), with the same degradation conditions as Group A, and using the same algorithm for traffic sign recognition. By measuring the average recognition accuracy rate of the two groups, we can see that the recognition accuracy rate of the suggested traffic sign is 82.5%, significantly higher than the 75.8% of the original traffic signs.

5.1.2 Distance based Accuracy Estimation vs. Real Classifier Accuracy

As noted during presentation, training with fixed features requires approximately 1 minute per iteration. Thus, 1000 iterations with simulated annealing is feasible within a day. We supplement experiments by training models using fixed MobileNet-v2 features, and compare our proposed distance-based accuracy estimation (Section 4.3) against using fixed features directly for accuracy. The resulting accuracy vs. blur curves are highly similar, validating our proposed approach. While real iterative training could provide greater accuracy, it incurs high cost (1000+ minutes). For an interactive system that guides traffic sign design via optimization, this training time is prohibitive.

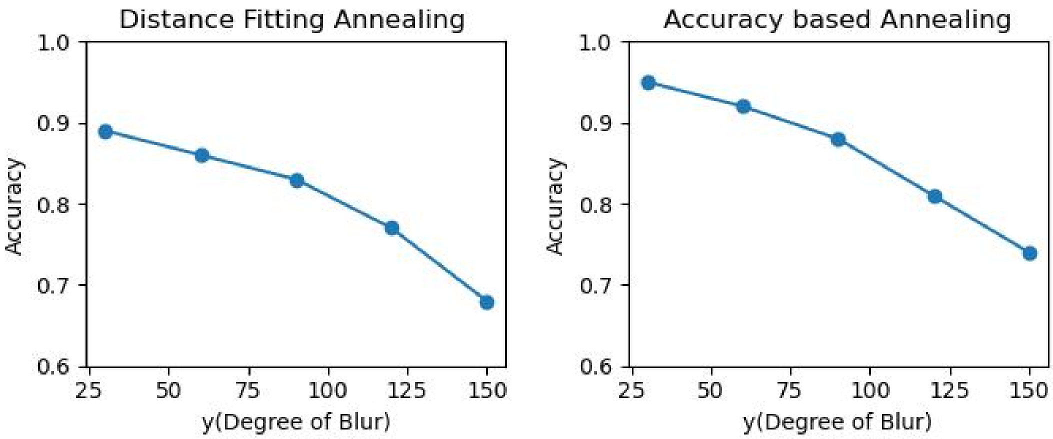


Figure 13 Results of fixed feature training and simulated annealing

5.2 Generalization to More Models for Blurry Traffic Sign Recognition

Our initial experiments only evaluated MobileNet-v2 due to time limitations. As noted, our approach is classifier-dependent, i.e. effectiveness on MobileNet may not transfer to other networks. We supplement additional experiments on Swin-Transformer, ResNet50, and VGG16 to validate: 1) whether our method improves design accuracy across models, and 2) whether designs optimized for one model improve others.
Modular ML frameworks allow easy switching between classifier architectures. Results show accuracy gains across models using our method, indicating general effectiveness. Interestingly, designs optimized for one model still improve others, suggesting features capture universal patterns. Detailed results are in the Experiments section.

```
net_mobe = get_model('mobilenet-v2_8xb32_in1k', '/content/

net_swin = get_model('swinv2-tiny-w8_3rdparty_in1k-256px',
net_res50 = get_model('resnetv1d50_8xb32_in1k', pretrainec
net_vgg16 = get_model('vgg16bn_8xb32_in1k', pretrained=Fal
```

Figure 14 Part of code of switch model

Model	mobileNet-v2	Swin-Transform	ResNet50	VGG16
slight blur, original road sign	97%	100%	91%	99%
slight blur, algorithm-selected road sign	98%	100%	99%	100%
accuracy rate increase	1%	0%	8%	1%
severe blur, original road sign	69%	90%	78%	86%
severe blur, algorithm-selected road sign	80%	97%	90%	90%
accuracy rate increase	11%	7%	12%	4%

Table 1 Result of accuracy rate of different classifier

For both slight blur (simulated degree of blur=150) and severe blur (simulated degree of blur=250), Swin Transformer achieved the highest accuracy rate with the original design. However, with our proposed design, the accuracy improved substantially for all classifiers, validating the effectiveness of our approach.

	mobileNet-v2	Swin-Transform	ResNet50	VGG16
mobileNet-v2	0.11	0.08	0.09	0.10
Swin-Transformer	0.08	0.07	0.08	0.08
ResNet50	0.12	0.13	0.12	0.14
VGG16	0.15	0.13	0.13	0.04

Table 2 Gain in recognition accuracy of other models after using different base models for design (post design accuracy - original accuracy)

We further conducted cross-validation by testing if design of model A improves model B. While the improvement is generally largest for model A itself, it also works for other models. Interestingly, VGG16 only improved by 4% itself but provided larger gains for other models. This may be due to randomness and does not affect the overall conclusion.

5.3 Control Variable Experiment for Sign Design Principles

Principle 1: Information should be distributed vertically rather than horizontally.

As the images of the signs are more likely to be subjected to horizontal blur when the car is in motion, such blur can harm the accuracy of sign recognition. If the information is distributed more horizontally, i.e., the variance of the image in the x-direction is larger, it is more prone to blur. Therefore, the information should be distributed vertically as much as possible, that is, the traffic signs should have a larger variance in the y-direction, as the information in the vertical direction is less susceptible to motion blur.

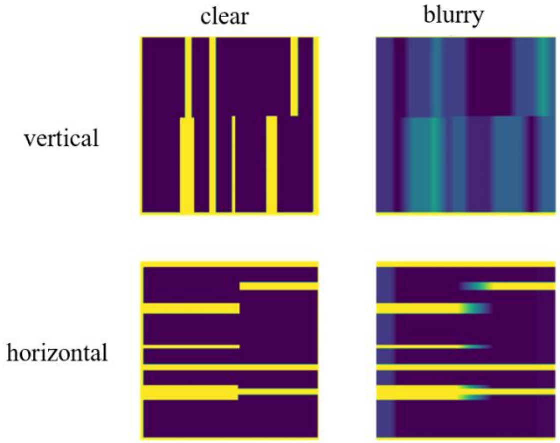


Figure 15 Classification Accuracy under Vertical and Horizontal Information

To validate our findings, we conducted an experiment using images with information oriented either vertically (horizontal stripes) or horizontally (vertical stripes). The classification accuracy was compared using fixed features. We introduced motion blur to the original images, with the degree of blur randomly chosen between 50 and 300. It can be observed that the distribution of horizontal stripes remains relatively clear even after blurring, while the vertical stripes become difficult to discern, resulting in a significant difference in accuracy. The accuracy of horizontal stripes reaches 99.1%, whereas that of vertical stripes is only 58.5%.

Principle 2: Diversified Outer Contour of Road Signs

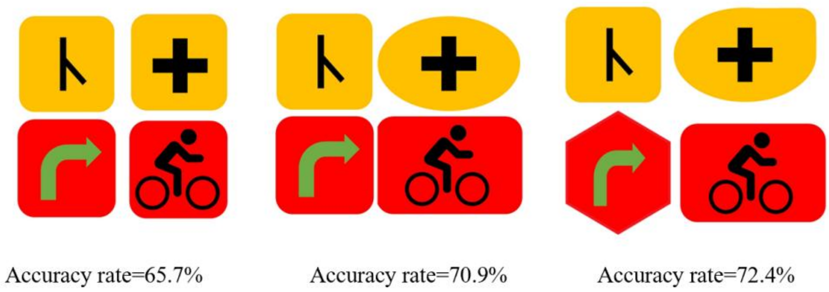


Figure 16 Classification Accuracy under Varying Contour Richness

We present examples from our experiments using three different scenarios: fixed shape, existing road signs, and enhanced outer contours. The accuracy of classification for existing road signs is 70.9%. When all outer contours of road signs are degraded to squares, the accuracy drops to 65.7%. However, by using more diverse outer contours, the accuracy of road sign recognition improves to 72.4%.

5.4 Verification of Experiments

Thus far, we have assumed the traffic signs are already localized and focused on classification accuracy.

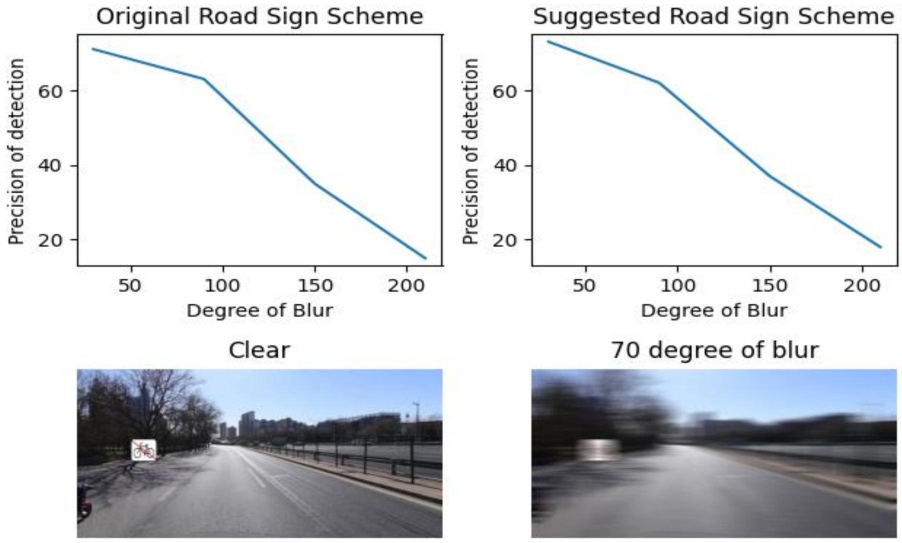


Figure 17 We paste the original and redesigned road signs into images for simulation testing experiments

To validate the impact on detection, we simulate detection experiments by overlaying original and redesigned signs onto real images.

After the province-level contest, we conducted additional experiments on object detection using YOLO, as many students have gained proficiency in this. We simulate detection by digitally adding existing and redesigned signs onto real background images and ask collaborators with detection expertise to run experiments.

The results show our redesigned signs also moderately improve detection model performance. This is likely because our signs have slightly higher visual diversity and salience compared to existing designs.

5.5 Preliminary Real-World Video Capture Experiments

Figure 18 Data collection under the controlled rail system

We conduct preliminary real-world video capture experiments using a controlled rail system and small sign models to validate the improved performance observed in simulations. Capturing real-world traffic signs from vehicles poses challenges - it is difficult to quickly access traffic signs in the experiment, and self-designed signs cannot be erected roadside. Our rail system enables reproducible capture approximating real conditions. As depicted in Figure 18, it comprises a driver, a motor, a belt track, a limiter, and a console. We control the movement of the console by controlling the motor operation with a driver, which in turn drives the belt track to rotate. The driver controls the direction and speed of operation, and with the help of the limiter, the operating range of console can be limited.

The maximum speed is 1.5m/s. We use a speed of ~0.7m/s for capture. Since signs are scaled down ~30×, this corresponds to 21m/s or 75.6km/h in a real vehicle, typical of urban driving. We print original and optimized sign designs onto wooden models for consistent capture.

Our procedure is: 1) Move the console via the controller. 2) Position sign models trackside. 3) Record video with a smartphone fixed to the console. 4) Extract some frames from the shooting results and use an open-source model for road sign detection to capture road sign images. 5) Compute recognition accuracy rate. Original signs have 78.6% accuracy rate, while our optimized versions achieve 85.7%, validating improved recognizability.

For better presentation, we construct a headworn display box from surface notebook and cardboard (Figure 4-c) and simulate motion blur programmatically. The specific implementation method is to obtain the current image and a series of images reduced by different multiples, and cut out the common parts for stacking. After stacking, real-time images with motion blur effect can be obtained under static conditions, and the difficulty of distinguishing different road signs under motion blur can be experienced by the human eye. This conveys the perceptual difficulty under blur, though without modeling distance effects. The simulation provides a visual sense of motion blur but does not correspond to a particular real-world speed and observation.

6 Conclusion

We propose redesigned traffic signs with added features that are easy to recognize and hard to blur, improving recognition accuracy in challenging conditions. Compared to electronic signs [9], our approach has advantages:First, electronic signs have reliability risks from power and hardware failures that would deprive autonomous systems of critical information. Second, digitization introduces vulnerabilities to network attacks, where humans cannot intuitively detect inconsistencies between electronic and physical signs, exacerbating safety issues. Beyond accuracy gains [10,11], security is an important consideration for digital sign design that we leave for future work. Nonetheless, our redesigned signs with more salient shapes and colors are inherently more robust to recognize.

In summary, we establish design principles for recognizable signs, generate optimized combinations via simulated annealing, and experimentally validate improved blur robustness over existing signs. This research has practical value for improving the safety of assisted driving, particularly as autonomous technologies become widespread.

At the same time, this article actually raises a new class of questions, namely, can the design of a fixed classifier/classification network be changed in an application to make classification easier to recognize. This article proposes the use of simulated annealing and provides a preliminary solution, which has been applied in the field of road sign recognition and expanded to the problem of animation avatar design. Such an algorithm that can design more discriminative categories may have broader applications in the field of design. An end-to-end learnable approach could produce superior results. The cross-model generalization suggests human effectiveness is worth investigating. Overall, the algorithmic category redesign capability may have broad utility for design.

References

[1] Toyota, road sign assist [EB/OL], (20230105). https://www.toyota.com.au/toyota-safety-sense/road-sign-assist .

[2] BMW,speed limit information, [EB/OL], (20230105) . http://content.bmwusa.com/microsite/3seriesActive Hybrid_2013/com/en/newvehicles/3series/sedan_active_hybrid/2011/showroom/safety/traffic_sign_recognition.html.

[3] AudiTraffic Sign Recognition[EB/OL], (20230105). https://www.audi.ca/ca/web/en/search-terms/Traffic-Sign-Recognition.html.

[4] Sermanet P , Lecun Y . Traffic sign recognition with multi-scale Convolutional Networks[C]// International Joint Conference on Neural Networks. IEEE, 2011:2809-2813.

[5] Zhe Z , Liang D , Zhang S , et al. Traffic-Sign Detection and Classification in the Wild[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[6] Wali S B , Abdullah M A , Hannan M A , et al. Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges[J]. Sensors, 2019, 19(9):2093.

[7] Kupyn O , Budzan V , Mykhailych M , et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[8] Zhu M , Han K , Zhang C , et al. Low-resolution Visual Recognition via Deep Feature Distillation[C]// ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

[9] Toh C K , Cano J C , Fernandez-Laguia C , et al. Wireless Digital Traffic Signs of the Future[J]. Iet Networks, 2018.

[10] Yuan Y , Xiong Z , Wang Q . An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition[J]. IEEE Tr ansactions on Intelligent Transportation Systems, 2017.

[11] Zhang J , Xie Z , Sun J , et al. A Cascaded R-CNN With Multiscale Attention and Imbalanced Samples for Traffic Sign Detection[J]. IEEE Access, 2020, PP(99):1-1.

Acknowledgements

My project went through a year-long process from choosing the topic to implementing and modifying it, followed by iterations. During this process, I received a lot of help from teachers and experts, especially at the stage of selecting the topic. Gao Kai, a teacher from Beijing No. 2 High School, provided guidance on some of my ideas and creativity, and together we finalized the topic for this project. Early in 2023, I was successfully admitted to the China Association for Science and Technology's and the Ministry of Education's high school student development program called "Talent Program". I had the opportunity to study at the State Key Laboratory of Complex Systems and Control, Chinese Academy of Sciences, Autonomous Research Institute. I would like to thank Dr. Liu Xiwei, Ms. Gong Xiaoyan, and Dr. Zhao Hongxia for their support provided in the project research and experiment process. My project won first prize in the 42nd Beijing Youth Science and Technology Innovation Competition. In the review process, many experts (whose names were anonymous) also provided many suggestions for improvement. Finally, I would like to express my gratitude to Li Cheng, a researcher at Sensetime Group, who encouraged me several times when my project encountered difficulties and kept me going.

Appendix: Application to Anime Character Design

Our algorithm is not limited to autonomous driving and can be applied to selecting a subsets from any candidate pool which has relatively better classification accuracy rate. As an example, we apply it to anime character design using StyleGAN-generated portraits.

Figure 19 Random and algorithm selected portraits

Comparing the random (left) versus optimized (right) selection of 55 portraits from 165 candidates, our simulated annealing approach visibly chooses more diverse faces. With visualization based on nearest neighbor in feature space, the left lacks variety and contains many similar pairs. Thus, even in this very different domain of anime design, our algorithm successfully identifies a subset with lower overlap, validating the general applicability.

We can see that our selection to have more facial feature diversity. In summary, while we focus on traffic signs, the algorithm proposed provides a novel general solution for selecting maximally diverse subsets from candidate pools. The anime experiment highlights the versatility beyond autonomous driving and suggests promise for creative applications like generative art.

为自动驾驶设计路牌？使用退火方法设计更易识别的路牌

学生 花弘扬 北京二中

指导教师 高凯 北京二中

摘 要

在高速驾驶的场景中，路牌识别面临着各种挑战，如运动模糊和低分辨率等。传统的研究者往往致力于提升识别网络的能力，然而会随之带来推断时间上升等问题。本文从另一个角度出发，提出了一个新的数学问题：即如何设计适应于复杂环境下，更容易被计算机识别的路牌体系。

进一步，本文将路牌设计问题简化为一个组合问题，即能否从一定的候选集中，选择出 M 个路牌，使得其对应的训练-测试错误率能够达到最小。并且围绕这样一个问题，结合了机器学习和模拟退火，给出了一套最优的路牌选择方案。该方法考虑了 M 组候选路牌，每组 K 个，即 KM 种路牌之间的相互关系，最终得到 M 个最适合的路牌，相互之间尽可能不混淆。实验结果显示，本文提出的路牌设计方案在各种模糊场景下都具有优良的识别准确率。

本文提出的路牌设计方案，可以为政府设计新的国标路牌提供参考建议。另外在特定的封闭环境中，如港口和矿场等，该方案也可以提高自动驾驶技术的可靠性。因此，在自动驾驶系统普及的时代，该方案可以被广泛应用，降低因路牌识别错误而导致的交通事故发生率。同时本文提出的新问题和对应的算法，也可以在更多的场景（如动漫人物设计等）中使用。

1 引言

根据美国交通部 2018 年的一项分析，94%的交通事故是由司机引起的。为了减少交通事故的发生，自动驾驶系统和辅助驾驶系统得到了迅速发展。很多车企开始更加注重使用自动驾驶和辅助驾驶相关技术，如本田、丰田、奔驰、宝马、奥迪等，都开始在新的车型中利用摄像头识别限速牌，并对车主进行提示或辅助限速[1-3]。

由于视觉信息是现有交通系统的主要组成部分，所有现有自动驾驶系统的关键功能都依赖于摄像头来为道路状况识别提供视觉信息，并配有激光雷达和雷达，用于防止汽车碰撞。在路况识别方面，道路标志的识别是需要解决的重要问题之一。

在机器学习相关的课程上，我们了解到机器学习需要通过数据集来对样本的分布进行学习，以此建立机器学习的模型，从而实现识别的目标。一般来说，人工智能研究者往往会致力于对现有的给定路牌，提升算法能力从而提高摄像头拍摄图片的识别准确率，比如丘成桐杯 21 年的银奖获得者 Richard Xue 的 CT 分类工作，就采用了多角度+多任务训练的方法来提升性能。然而，这么做仍然会遇到很多问题：1）很多路牌从设计上就是容易混淆的，存在很多看似接近但含义相差巨大的路牌，很容易造成混淆；2）很多路牌图案设计过于复杂，从远处看的关键信息不是很清晰；3）现有缓解低像素或者抗模糊的识别办法，大多依赖深度学习网络，实时性不高，一定程度上降低了安全性；4）现有基于深度学习的路牌识别系统大多基于现有数据训练生成，在遇到没有学习过的新款路牌或特种路牌时，迁移识别能力不强，需要重新训练来获得好的识别模型，代价较大。

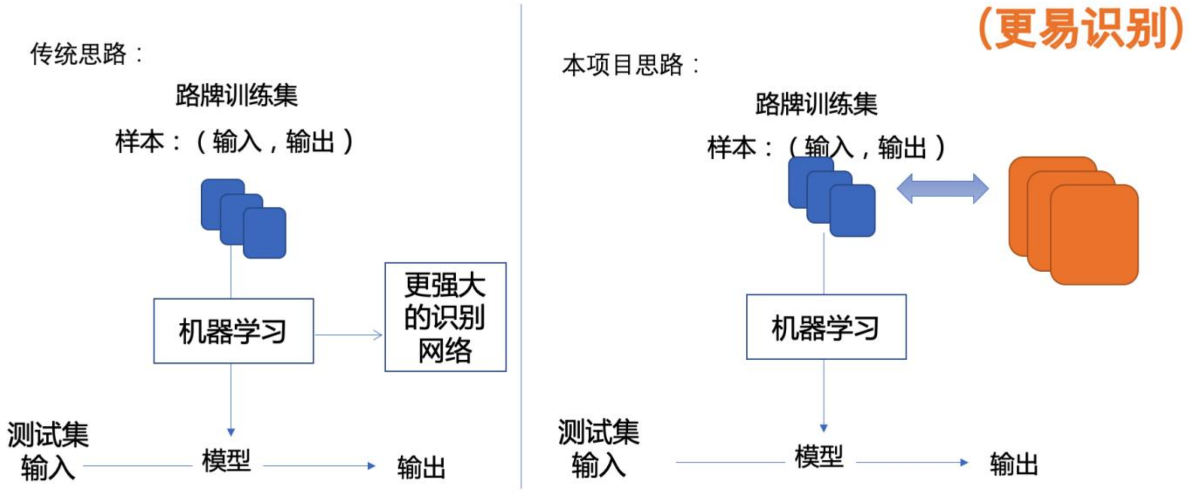


图 1 传统和本项目路牌识别的思路

于是我和指导老师讨论，能否去改变路牌识别的路牌，来提升整个系统的准确率。指导老师觉得尽管对于无人驾驶公司无法使用这个策略，但对于政府来说，有可能通过修订新的路牌，更换整套更容易被机器识别的路牌，来提升社会整体的无人驾驶和辅助驾驶的准确率。在创新大赛的省赛中，也进一步有专家指出，对于一个封闭的内部环境，比如矿场、港口等，也可以使用我们的方案来设计路标。需要进一步说明的是，本项目实际上提出了一种新颖的算法问题和初步的解决方案，这种设计、筛选训练标签的方法，也可以应用到除自动驾驶外的其他领域。在章节 5.6 中，我们给出了一个动漫人物设计的例子。

相比于仅仅去提升识别路牌的准确率，本文希望从数据的源头解决问题，即通过**提高路牌本身设计的合理性**，设计更易识别的整体路牌方案，来提高汽车无人驾驶系统的识别能力。我们希望能够通过一系列的研究，找到一套系统性的路牌设计方法，并且能够依据此思路**设计一套路牌，使得无人驾驶和辅助驾驶系统对路牌的识别准确率得到明显提升**，从而提升智能驾驶整体的安全性。

1.1 路牌设计的挑战

在 2023 年春天,最前沿的生成模型的技术正在快速发展。[3]比如 Stable Diffusion,可以输入若干句子描述,并生成相关的图像。这里我们也尝试用 Stable Diffusion 去生成与路牌相关的图像。我们编写了 30 句和路牌有关的句子输入到 Stable Diffusion 上,期待生成一些路牌图像。结果表明,Stable Diffusion 生成的路牌图像中,完全正确的设计较少,更多时候无法理解路牌的语义。即使算法可以生成正确的路牌图像,也很难保证生成的不同路牌之间不会产生混淆。直接依靠现有的生成模型自主设计容易识别的路牌组合,目前来说还存在一定困难。

[3] 在 2023 年春天的英才计划课上,老师向我们介绍了 Stable Diffusion 这种生成模型。我们在 B 站也看到网友用它生成各种图像。为了自己试试,我们编写了 30 句关于路牌的描述,输入到 Stable Diffusion 中,希望它能生成一些路牌图像。结果发现,有时它可以生成有趣的路牌设计,但更多时候完全无法理解语义。即使它生成了正确的路牌,也很难保证不同路牌之间不会混淆。直接修改 Stable Diffusion 的算法对我们高中生来说过于复杂。

			
A road sign indicate that parking car is restricted here.	A road sign indicating no bicycles allowed ahead.	A road sign showing no motorcycle.	A road sign showing there's a circular road.

图 2 Stable Diffusion 生成路牌的结果

1.2 转化生成问题为选择问题

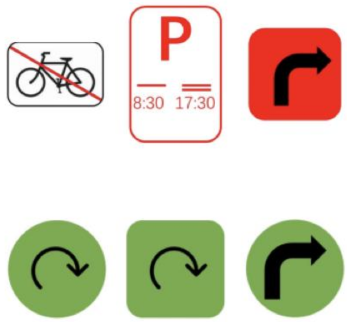
		
数据集中原来的路牌	添加的候选路牌示意	算法最终选择的路牌

图 3 路牌选择示意图

信息学竞赛中，往往会碰到很多组合优化问题，比如卡车可以选择不同的节点去运货，或者不同国家可以派遣不同的使者去外交，并且最终求出一个总价值最优的方案。所以我们是不是也可以考虑将我们的路牌生成问题，转化为一个优化问题呢？

如图 3 所示，假设我们预先设计了一套候选的路牌方案，我们假设 M 是路牌种类，并且我们为每种路牌设计了 K 个候选路牌，K 是每一种路牌方案的个数，那么所有路牌的候选图总数就是 KM 个，有 K^M 个可能情况（M 种路牌，每种路牌都选出一个方案）。于是，我们就可以将如何生成易识别的路牌的问题转化为一个在 K^M 个的设计路牌组合中找出分类准确率最高的组合的问题。

这样，路牌设计就被转化成了人工设计+最优选择的问题。要在给出的组合中找出准确率最高的一组，也是很复杂的，需要有一种算法来做到快速的最优组合搜索。本文选用了模拟退火算法（Simulate Anneal Arithmetic，SAA），搜索到识别效果理想的路牌方案，并通过实验验证了所搜索路牌方案的效果。

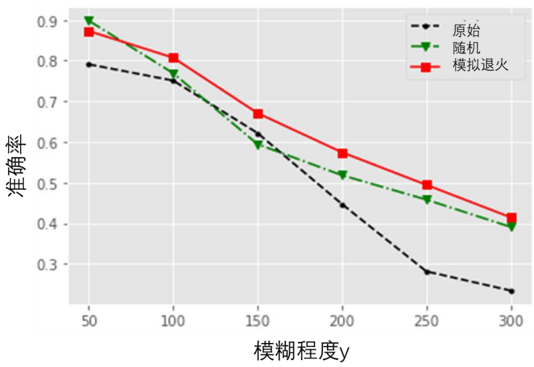


图 4-a 不同模糊程度下识别准确率

横轴是路牌不同的模糊程度，纵轴是模型在几个方案下的识别准确率，可以看到我们的模拟退火方案在高模糊下取得了更好的识别准确率。

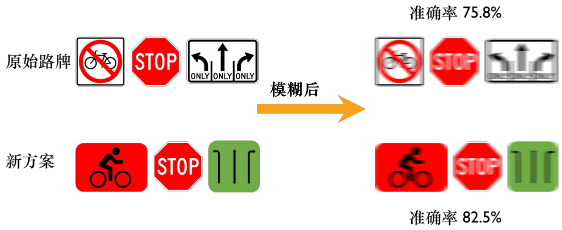


图 4-b 模糊程度 100 情况下原始路牌和新方案的比较

在各种模糊条件下，新方案比原始路牌方案平均提升了12%的准确率。

图 4-c 头显盒子示意图

受到 MIT 一个工作现场展示方式的启发，我制作了头显盒子，戴上盒子之后可以体验路牌相对于摄像头在不同位置时的模糊情况。由于人眼的成像原理与摄像头有所不同，实际上摄像头在高速情况下获得的图片往往比人眼捕捉到的更模糊。

论文的主要贡献是，路牌重新设计选择以后，路牌的分类准确率大大提升，可以降低无人驾驶/辅助驾驶因为路牌识别错误而导致的交通事故发生概率。具体而言，本文创新点包含如下几个方面：

1. 提供和设计了一套分类准确率更高的路牌方案，并设计了一种基于模拟退火算法的最优路牌组合遴选方法。

2. 打破了传统的识别优化思路，为以后的智能识别提供了新思路：不改变识别系统，而是通过改变被识别物来提高识别准确率。最终实验表明，在不同的模糊程度下，我们设计的算法相比于随机选取的路牌或者现有方案，能够在 55 种路牌分类上，不同模糊程度下提升 6-20%的识别准确率。

3. 为了验证我们方案设计、筛选出的路牌方案有普遍性，我们增加了跨模型验证方案，验证了由 A 模型设计方案后，B 模型的识别结果仍然会有显著的提升。验证了我们方案的普遍性，为其他类似研究工作提供一种可行的方案，指导其设计相应的最优策略选择方法。

本文的剩余部分按照以下方式组织：在第三章我们会测试深度学习模型，包括分类模型和检测模型在运动模糊下的性能下降情况。在第四章介绍我们的路牌设计算法。在第五章我们会进行大量的实验，来验证我们方法的有效性和泛化能力。

2 研究现状

2.1 无人驾驶中的视觉检测与识别

无人驾驶大量运用了路牌、车辆自动检测与识别技术。其中，在路牌识别方面，模糊和远处低分辨率路牌的识别一直是个很有挑战的问题。一些文章指出，路牌的坏损、低分辨率、运动模糊等是识别错误的主要原因[4-6]。其中路牌的坏损很难通过其它技术手段来解决，所以本文重点考虑低分辨率与运动模糊下的路牌识别。

2.2 模糊与低分辨率图片识别

模糊和低分辨率是图片常见的坏损形式，主要由于距离，对焦模糊和运动等原因产生。低分辨率图片的识别，有两种思路：一种是进行图片的分辨率恢复，恢复后再识别[7]；另一种是直接针对模糊图片进行训练识别[8]。但两种方法都只能在一定程度上缓解低像素带来的识别准确率下降问题，随着模糊的加剧，识别准确率依然会迅速下降。

2.3 路牌改动

有车企提出改善路牌材料，提升夜间的可视度。另外，还有研究建议使用电子路牌发送信息来和车辆进行互动，帮助车载系统识别路牌[9]。但由于成本等原因，类似方法都没得到广泛的应用。利用现有的 AI 生成方法，也可以进行路牌的生成，但效果并不好，一些结果如图 2 所示，路牌特别是人眼识别起来的可读性较差。

3 运动模糊路牌识别的问题与挑战

对一个基本的识别系统，当在受到运动模糊等其他的因素的干扰时，效果就会下降。在这一部分，本文将通过实验验证，即使使用了一些先进的深度学习方法，识别准确率仍然会产生大幅度下降。

3.1 对实拍照片的定性观测

本文在实际生活中收集了一些实验数据，具体方法如下：坐在副驾驶位置使用苹果 11（1200 万像素）拍摄途中遇到的两个限速路牌，如图 5 所示。用人眼可以差不多看出右边路牌是限速 30（其实也有些像 20），此时机器也不一定能准确进行识别，分类准确率可想而知。而左侧的路牌，可以说是一片模糊，完全看不清。

可以考虑如下的问题：右侧路牌到车运动方向的所在纵向距离大概是 20 米（沿公路方向的距离），车到路牌的所在路边横向距离大概是 10 米，车速大概是 50km/h。那么，司机的反应时间只有 0.5 秒，对应于行驶七米的距离。对于一个正在以 14m/s 运动的汽车来说，指望车能在剩余的距离内就做出反应也有些难度，十分容易发生车祸。即使是相对清晰的图中右侧路牌，在摄像头拍照下也并不算清晰，如果识别错误，误识别成 80 了，那在自动驾驶或需要车辆进行辅助分析的情况下就更容易发生车祸了。

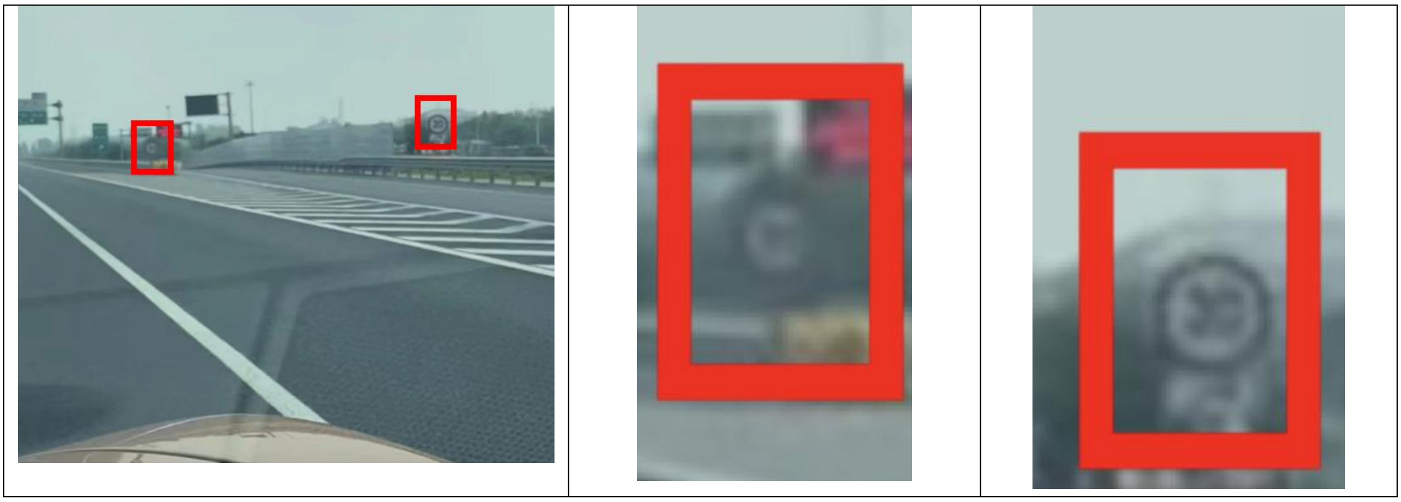


图 5 公路上实际拍摄路牌的情况

3.2 清晰路牌识别基准算法

首先需要对清晰路牌进行识别，作为本文研究的一个基准，同时也可以验证分类方法的有效性。在这里我们使用 MobileNet_v2 在一个 55 分类的交通路牌数据集上进行初步的训练和验证。路牌来自于 Lisa 交通路牌数据集，这是一个美国路牌的数据集。

由于数据量有限，对路牌图像还需要进行数据增广。考虑到路牌的特殊性，无法进行翻转操作，这里的增广主要涉及亮度、小角度的选择以及小幅度的仿射变换。每种路牌中分出 20%用于构建测试集。本文采用的预训练模型是 MobileNet_v2，这种模型是针对移动端设计的小模型，在训练速度较快的情况下和 Resnet50 相比仍然可以保持较高的准确率，本文中端到端训练也是基于这种模型架构。

通过测试得到结果，直接抽取特征得到的准确率为 76.3%，微调训练后的准确率上升到 87.8%。由此可见微调训练后模型识别路牌的准确率更高。下面的实验中也会以微调训练的模型作为最终的标准。同时，如图 6 所示，可以利用 t-SNE 进行特征空间的可视化，t-SNE 是一种数据可视化的工具，能够将高维数据降到 2-3 维（降维），然后画成图，并尽量保持特征之间的距离。t-SNE 结果显示，端到端训练后路牌在特征空间中分布更加分散，即使模糊后不同路牌间也可以有较明显的空间分布上面的差别，由此说明微调训练确实可以提升路牌识别的准确率。

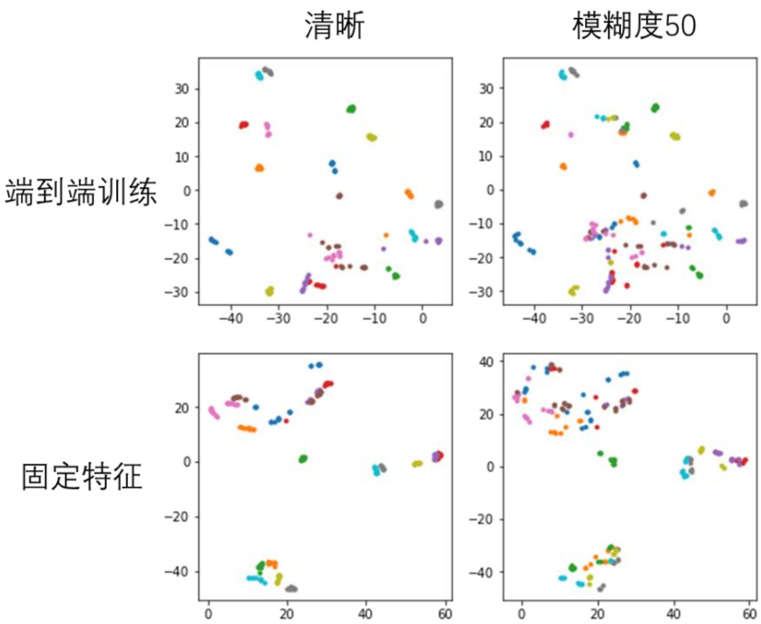


图 6 利用已有模型以及微调后不同模糊程度下 t-SNE 可视化结果

路牌识别准确率不高是有多方面原因的，一方面清晰路牌在旋转等情况下依然会发生较大的变化，同时由于网络仅仅是微调训练，数据量有限，特别是一些路牌仅仅有数字的区别，而本文采用的预训练图片分类模型识别的重点并不在此，所以会对相似路牌产生误识别的情况。在本文中我们会先使用 MobileNet_v2 进行实验，因为本文重点讨论的是路牌的设计，所以讨论都是基于 MobileNet 模型。对模型结构进行提升的思路和本文的设计也并不矛盾，如果有更利于识别的路牌系统，系统准确率的提升也会更容易。当然，在创新大赛区域赛事上，很多评委向我指出了单个模型训练会导致方案和结论的泛化能力缺失的问题。所以在后续的交叉对比中，我们会提供 VGG16,ResNet50 和 Swin-Transformer 的实验结果，以加强我们的论证。

3.3 模糊路牌识别

为了模拟真实的复杂场景下路牌识别，主要是需要考虑不同的坏损操作，操作算符记为 O，并将多种操作结合。真实复杂场景下，图像之所以会模糊是因为图像受到了干扰。我们所说的干扰包括运动模糊、对焦模糊、低环境照度、远距离低分辨率以及图像噪点。由于车在不断的运动，并且照相机的快门速度不是无限快，是有时间跨度的；快门开启的这段时间,物体由于快速运动,会走一段距离,物体的像也会在底片或者图像传感器上移动一个距离，这样许多像叠在一起，就导致了相片模糊不清。

对于函数 O，本文使用了四种不同的算法来模拟四种不同的模糊效果，这四种模糊效果与四种可能影响交通标志识别精度的实际情况相关：加速度导致的运动模糊、远距离导致的低分辨率、散焦导致的高斯模糊以及较低环境照度引起的高噪声。图 7-a 展示了两个常见的交通标志（限速标志和调头标志）的例子，并使用我们的算法模拟摄像机在这些情况下可能获得的图像。

对四种坏损，其中每种坏损操作都有对应的坏损系数。运动模糊中模糊发生在水平方向正负 10 度的范围内，坏损程度由模糊核的大小 y_m 表示；低分辨率情况下，当坏损程度为 y_l 时，输出图片分辨率为 $40/y_l \times 40/y_l$。高斯模糊中，以坏损程度 y_g 作为高斯核的大小。增加噪声中采用的是高斯噪声，将生成的高斯噪声图以坏损程度 y_n 的倍数附加在原始图像上。

图 7-a 不同的坏损模拟示意图

图 7-b 高速公路实拍路牌

3.4 模糊路牌对于识别的影响

如前面所述，在各种模糊的情况下，路牌识别的准确率会随坏损程度的加剧而逐渐下降，在这里通过实验对此进行验证。实验采用了从 50 到 300 不同模糊程度的图片运动模糊。在固定特征和端到端训练模型下，识别准确率均发生较大程度下降。这里我们定量给出不同模糊程度下，分类和检测的准确率下降情况。

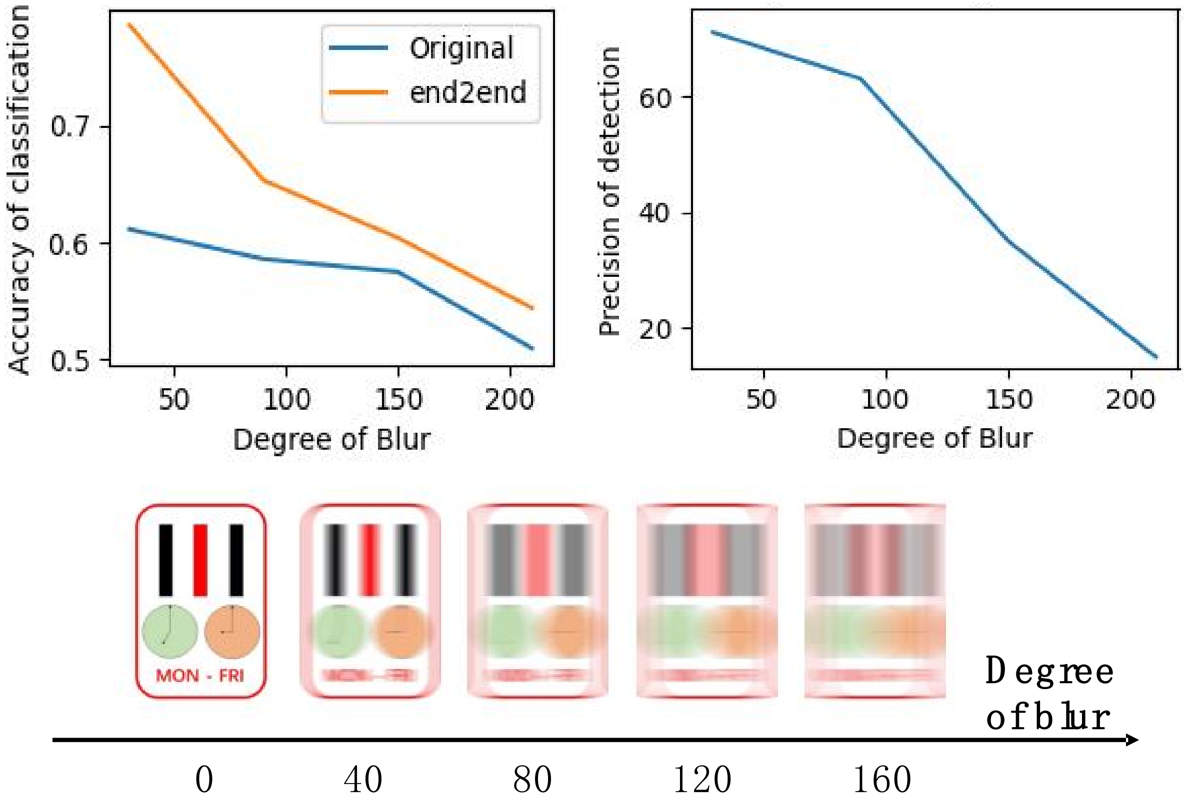


图 8 无论是端到端训练还是固定特征，在模糊的情况下都会出现严重的准确率下降，同时目标检测的性能也会出现严重的下降

可以看到在比较严重的模糊下，无论是分类模型还是检测模型都会有严重的性能下降。即使很多工作可以通过加强网络的方式，去提升性能，但是这些方法往往会增加推断所需要的时间，并且与本文的方法不冲突，可以结合使用。在这里我们重申课题的目标：我们希望找到一个更合适的路牌设计方案，使得在不改变模型或者训练方案的情况下，提升路牌系统在不同模糊程度下的准确率。据我们所知，这是一个比较新颖的问题，之前较少有人研究。

4 路牌设计与识别

4.1 核心框架

我们在章节 1.2 中提到了，我们把将如何生成易识别的一套路牌的问题转化为一个在 K^M 个路牌组合方案中找出理想方案的问题。这里我们考虑 K=3，M=55 的情况。理想方案指在存在模糊的情况下这一组路牌设计方案的分类准确率能达到最高。具体设计中对每种路牌都给出了两个可供替换的方案。

显然，最准确的方法是对于选定的一个 55 种路牌的方案，重新抽取路牌数据的训练集进行训练，然后对应在测试集上进行测试。不过这样即使是使用很简单的线性分类器，也需要很长的时间。我们必须找到一种算法，这种算法可以：1）快速评估一个路牌方案的准确率；2）大幅压缩 K^M 的搜索空间，减少搜索的次数。

考虑到路牌的方案很多，选择量会巨大，即使在路牌种类较少如 M=10 的情况下，方案就有接近六万种可能了，依次去进行验证找到准确率最高的方案将花费大量的时间。所以本文选择使用一种适用于在一个大的搜寻空间内找寻最优解的算法——模拟退火算法，来近似获取理想方案。

4.2 分类准确率与特征空间中的数据分布

本项目中如限速路牌的数字，左右转弯箭头朝向就是路牌之间进行区分的特征（当然也包括颜色和形状，如圆的，三角的，蓝的，黄的等）。而特征空间里的数据点，就是将路牌图片的特征由神经网络数字化后得到的。神经网络对图片的分类就是基于特征进行的，当特征空间上两种路牌对应的特征点区域存在重叠时，分类器并不能将两者完全分开，从而产生了分类错误。可以认为当两种路牌对应的特征点在特征空间中重叠越多时，分类的错误率也就越高。

图 9 能够更直观的展示这一现象，假设有 A、B、C 三种路牌，每种路牌有 K=2 种方案。每种路牌的重合部分越小，也就意味着这几个方案的路牌的特征越容易被区分开，即机器更容易区分这些路牌，分类准确率就高。但如果重合部分大，就说明这几种方案的特征更相似，机器在做分类的话，更容易出现混淆，就比如 9 图中的方案 A2、C1 和 B2，它们的重合部分多，所以当机器分类的时候，分类准确率就会低。它不能确定识别的到底是 A1 还是 A2，因为有些特征 A1、A2 都具备。那么，对于这种图像分布，机器的分类准确率就低。而 A1、C2、B2，三个方案就没重合，分类准确率就高。设计路牌也是这个道理，我们需要尽可能设计出，在特征空间内重合部分更小的路牌方案，即区分度更高的路牌。重叠越少，方案越理想。

理想方案是想要路牌分类准确率越高越好，那么反映在特征空间上就是让任意两种路牌之间的类中心的距离尽可能远，两个圆的重叠面积尽可能小。这里可以采用 t-SNE 方法进行可视化。当抽取特征并用 t-SNE 特征空间可视化后，我们希望新的设计方案可以使每种路牌的候选方案距离不要太近，离得越远，也就代表路牌越容易被区分。

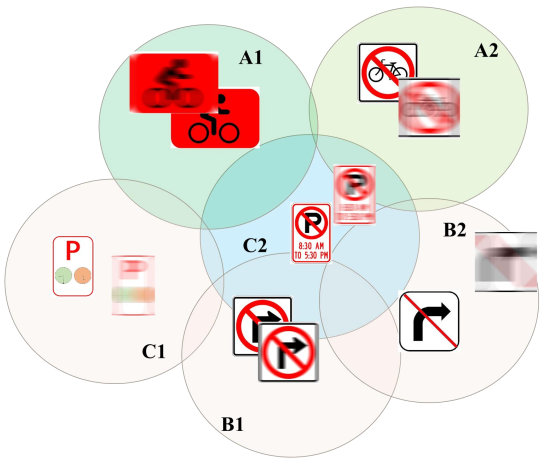


图 9 特征分布示意图

现在面临的问题就是，怎样用算法找到类中心距离尽可能远的理想方案。这里要强调一下，对任意图片提取到的特征是等长的，不会出现由于运动模糊让特征数字变少这么一说。定义 y 为图像的受损程度。y 小，也就意味着图像受损的程度小。当物体图像受到噪音等干扰时，提取特征将会是模糊的物体特征。把模糊作为一种干扰，这种干扰会令不同路牌的特征分布范围变得更大。如果 y 值越大的话，图像的特征受到的干扰也就越大，使得提取的特征更远离清晰图像的特征，也就是圆形的中心，进而使得特征空间里的数据点（特征点）分布得更缺乏区分度。

模拟退火算法是一种通用概率演算法，用来在一个大的搜寻空间内找寻命题的最优解。模拟退火来自冶金学的专有名词退火。退火是将材料加热后再经特定速率冷却，目的是增大晶粒的体积，并且减少晶格中的缺陷。材料中的原子原来会停留在使内能有局部最小值的位置，加热使能量变大，原子会离开原来位置，而随机在其他位置中移动。退火冷却时速度较慢，使得原子有较多可能可以找到内能比原先更低的位置。

模拟退火就模拟了这个物理过程，通过定义能量函数 E，并通过“不断尝试”小幅度修改当前状态，伴随着温度 T 降低的过程，在随机移动的过程中逐渐向 E 更低的状态进行选择，并最终找到最优解。以本文考虑的路牌为例，上面我们有提到有 3^{55} 种方案，模拟退火算法会先随便选一个情况，不断改变里面的路牌方案，并且尽量保留尝试出来的好的方

案，并在此基础上继续不断尝试。虽然有的时候会试出不好的方案，但是在不断的尝试下，模拟退火算出来的方案会逐渐的越来越好。

模拟退火需要在给出一组方案后对方案的好坏进行评价，这里如果以准确率进行评价，仍然需要进行一组测试，花费较多时间，所以考虑近似地以特征空间上各个类别之间的距离来近似得到准确率作为评价标准。

4.3 距离-准确率关系

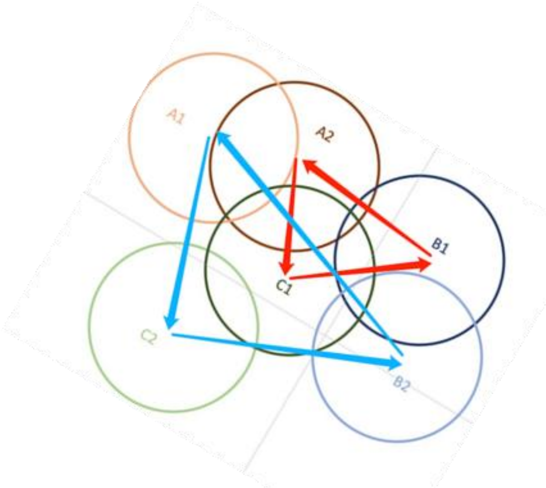


图 10 距离 d 示意图

模拟退火搜索的目标函数是能量 E，这里我们希望 E 可以直接对应于分类准确率。首先定义各个图片在特征空间的距离 d（两个方案的类中心连线距离，如图 10 所示的箭头）。我们设计一个公式，让 E 与我们的类中心距离 d 能符合特定关系：E 越小，d 越大。这样我们就可以进行拟合，近似得到准确率与所有类间距离 d 之间的关系。

为了获得更小的错误率P_e，d 越大越好，同时模拟退火冷却过程中能量 E 是越来越小的。所以取 E 为错误率P_e。考虑到 d 为非负数，所以我们让 d 成为分母（分母越大，分子不变，数变小）拟合式子是这样的，错误率P_e满足：

$$P_e = a/d^b . \qquad (1)$$

其中 a，b 为待拟合的参数，拟合结果如图 11 所示。

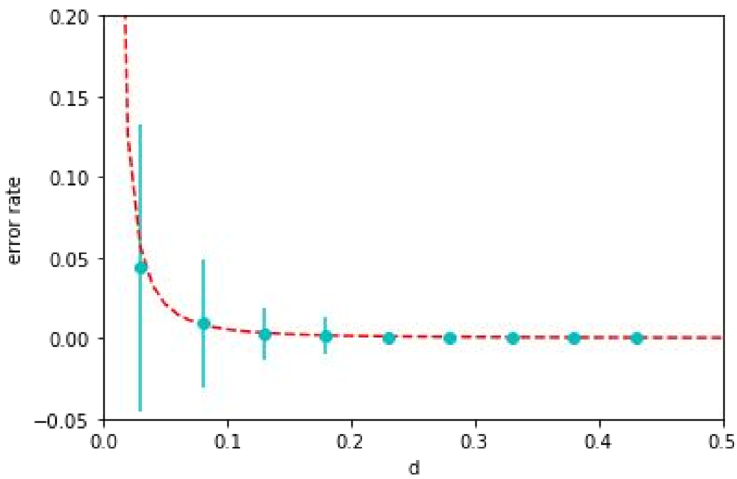


图 11 距离与错误率拟合曲线

由拟合结果可得能量 E 满足：

$$E = \sum_{ij} a/d_{ij}^{b}. \qquad (2)$$

4.4 模拟退火的实现

在用模拟退火算法寻找合理方案的过程中，会随机改变一种能量为 E 的路牌方案组合中某一个路牌的选择，并得到一个新的能量为 E_1 路牌组合，以下分两种情况进行处理。第一种情况：E_1 小于等于 E，保留 E_1 对应的路牌组合并继续随机改变路牌方案进行尝试；第二种情况：E_1 大于 E，则以一定概率选 E_1，选择概率 P 在下面给出。

需要强调的是，模拟退火算法中温度的定义尤其重要，若是温度很高的话，机器会不断的乱试，好的方案和坏的方案都会选择，而如果温度小的话，机器会一直选择 E 越小的方案，导致陷入局域最小。因此，我们会定义一个温度 T（温度随时间降低）来避免陷入局部最小。模拟退火算法中通常采用公式：

$$P = e^{-\Delta E/T}, \qquad (3)$$

随着时间 T 减小，P 也会随之减小，也就是说，随着时间的推移，遇到第二种情况选择新的 E 的可能会越来越小，也就是说保守选择好方案的可能变大了，这使得在前期不断尝试改变方案，到后期尝试越来越选择好的方案。可以理解为前期把可能情况全尝试，后面变成选尝试之后的好的方案，这样即避免陷入局域最小，还会让我们得到比以前更好的方案。

4.5 使用固定特征的结果代替距离估计[4]

模拟退火算法的迭代次数约为 1000 多次。固定特征训练的时间如果只需要 1 分钟,那么在有限的时间内(几个小时),算法也可以收敛得到结果。所以对于部分模型,可以补充采用固定特征训练,直接得到较高的准确率。这一点在实验环节已进行了汇报。

5 实验验证

实验章节将这样组织: 在 5.1 节我们会展示初步的结果，使用 MobileNet 和最简单的退火算法进行实验，并对比使用距离去估计准确率和直接使用端到端的真实准确率进行退火的性能差异。 在 5.2，针对很多评委指出的泛化能力的问题，我们会进一步补充 vgg, swin-transformer 等网络的交叉验证结果，证明我们方法的普适性。 5.3 是一个补充实验，来验证模糊方向对于识别准确率和路牌设计原则的影响。 5.4 是一个补充实验，因为项目大多的实验集中于图像分类方面，而模糊同样对图像检测任务也会有影响，所以在这里我们补充图像检测实验，验证我们设计的路牌是否在检测任务下也表现的更好。 在 5.5 中，我们尝试使用滑轨拍摄更真实的数据，并进一步验证我们的方案。

5.1 模拟实验

5.1.1 模拟退火的初步验证

在 55 组路牌中，每种路牌都绘制了两种备选图案，和原始路牌组成了 3 个选项。首先在清晰情况下提取每个路牌的中心点，用模拟退火找到最优方案。所有路牌在特征空间中的分布可以通过 t-SNE 算法在二维平面进行展示。如图 12 中左侧灰色点所示，一些路牌图案更加接近，在特征空间中也成团分布。模拟退火过程中能量 E 下降的过程如图 12 右侧所示，在模拟退火初期，温度 T 较高，能量 E 会在相邻轮次出现大幅度上升，随着温度 T 下降，概率 P 会随之下降。随着时间的增加，选择最近一次步骤的概率会逐渐变小，从一开始的疯狂随便尝试逐渐变得稳定。设定温度给了寻

[4] 在创新大赛区域赛的答辩中,一位评委指出模拟退火算法的迭代次数大概在 1000 次左右。考虑到如果用固定特征训练只需要 1 分钟,在我们有限的准备时间内(几个小时),算法也可以得到结果。所以我们对一些增加了使用固定特征直接训练,来得到模型准确率的实验。

找更优解的一个缓冲的机会，延长了寻找解的时间，这样使得最后我们会得到一个比刚开始就陷入局部最小的更好的解。此时选取的方案经 t-SNE 可视化后如图 12 左侧红色点所示，选中的各个路牌图案间都尽量保持了一定的距离。

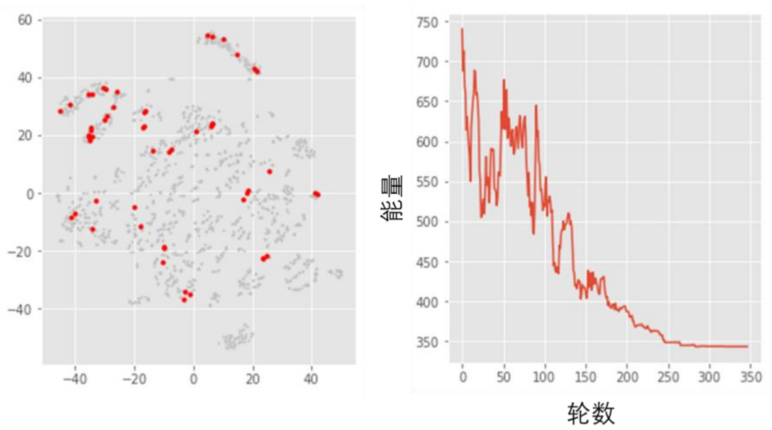


图 12 模拟退火结果

接下来可以对模拟退火算法给出的路牌组合进行准确率测试。在不同坏损程度下，对原始路牌、模拟退火路牌组合以及随机选定的路牌组合进行了准确率的测试结果如图 4-a 所示。

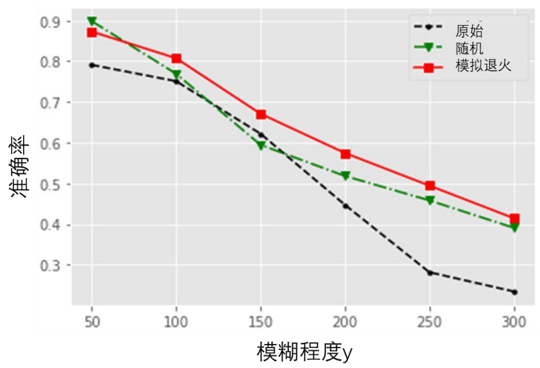


图 4-a 不同模糊程度下识别准确率

图中横轴为图像受损程度，纵轴是分类准确率。黑线是原本的图片，绿线是任意选择的路牌方案模糊后的图片识别准确率，红线是用模拟退火算法下选择的路牌组合模糊后得到的图片的准确率。对比红线和黑线，模拟退火后，分类准确率相比较于原始方案，在各个清晰度下均有明显提高，准确率提高近百分之十，随着模糊程度变大，分类准确率较原先提高了近百分之二十。观察绿线和黑线，随机结果的正确率在大部分清晰度下也低于模拟退火给出的结果，说明模拟退火有效地给出了一组更易于识别的路牌组合。通过分析我们可以知道，新设计的路牌对准确率提升具有一定效果，同时使用模拟退火算法也是十分的重要。

两种方案的一些例子如图 4-b 所示，下面新方案给出的路牌在模糊情况下具有更高的辨识度，互相之间的区分也更加明显，准确率更高。

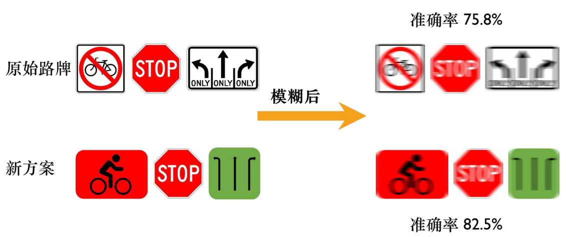


图 4-b 模糊程度 100 情况下原始路牌和新方案的比较

在各种模糊条件下，新方案比原始路牌方案平均提升了 12%的准确率。

原路牌的分类准确率在运动模糊、远处观测的条件下，会有显著的降低，而我们设计的路牌在相同条件下会减小运动模糊带给路牌识别的影响，进而提升分类准确率。如图 4-b 所示，最终实验结果表明，我们提出的交通标志在几种不同的模糊条件下在识别率上有明显优势。在本实验中，A 组由一组存在的交通标志组成，其坏损条件包括运动模糊、对焦模糊、低环境照度、远距离低分辨率以及图像噪点。在我们设计两组被选方案后，B 组是三种方案（设计了两组，原方案 A 是一组，2+1=3）中选取的最优结果，其坏损条件与 A 组相同，采用相同的算法进行交通标志的识别。然后通过测量两组的平均识别准确率可以看到，我们建议的交通标志识别准确率为 82.5%，远高于原始交通标志的 75.8%。

5.1.2 使用真实分类器的准确率进行退火

如果使用固定特征进行训练,每次训练时间约为 1 分钟。在模拟退火算法中,1000 次迭代的运行时间可以接受(小于一天)。这里补充了使用固定特征训练的实验。

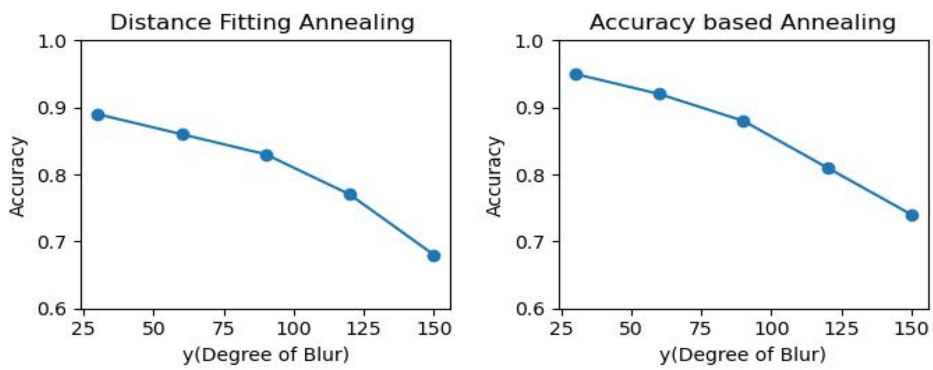


图 13 固定特征训练和模拟退火的结果

这里我们仍然以 mobileNet-v2 为基准特征，分别运行了本文提出的用距离估计准确率(章节 4.3)和直接使用固定特征训练来计算准确率的方法，并且对两者的结果分别测试了模糊度 vs 准确率的曲线。可以看见，两者的结果非常接近。说明我们在 4.3 提出的使用距离估计准确率的方法是有效的。对比来说，后者使用真实训练来估计准确率的方法可能更准，但是消耗的时间也更多。如果我们希望得到一个真实能够指导路牌设计的系统，并且能够迭代优化，1000 分钟的训练时间有一些过长了。

5.2 更多模型的测试与方案的泛化验证

本研究初期仅进行了基于 mobileNet-v2 的实验。为验证所提方法的通用性,增加了在其他分类模型上的实验:1. 所提方法是否在其他模型上也能提升路牌设计的准确率。2. 用一种模型设计的路牌方案,是否对其他模型也能提升准确率。

受益于通用的机器学习框架,可以简单替换分类模型。在这个实验中,补充测试了 SwinTransformer、ResNet50 和 VGG16。

```
net_mobe = get_model('mobilenet-v2_8xb32_in1k', '/content/

net_swin = get_model('swinv2-tiny-w8_3rdparty_in1k-256px',
net_res50 = get_model('resnetv1d50_8xb32_in1k', pretrained
net_vgg16 = get_model('vgg16bn_8xb32_in1k', pretrained=Fal
```

图 14 部分代码展示

模型	mobileNet-v2	Swin-Transform	ResNet50	VGG16
轻微模糊，原路牌方案	97%	100%	91%	99%
轻微模糊，我们的方案	98%	100%	99%	100%
准确率提升	1%	0%	8%	1%
严重模糊，原路牌方案	69%	90%	78%	86%
严重模糊，我们的方案	80%	97%	90%	90%
准确率提升	11%	7%	12%	4%

表 1 不同分类器的准确率结果

可以看到，在轻微（模拟模糊度=150）和严重（模拟模糊度=250）情况下，如果使用原来的路牌方案，Swin-Transformer 的准确率是最高的。 但是使用我们的方案进行设计后，各分类器准确率均有明显的提升。说明用我们的方案进行路牌设计是切实有效的。

	mobileNet-v2	Swin-Transform	ResNet50	VGG16
mobileNet-v2	0.11	0.08	0.09	0.10
Swin-Transformer	0.08	0.07	0.08	0.08
ResNet50	0.12	0.13	0.12	0.14
VGG16	0.15	0.13	0.13	0.04

表 2 使用不同基模型进行设计后，其他模型识别准确率的增益(设计后准确率 - 原准确率)

这里我们进一步去进行交叉验证。使用 A 模型设计路牌后，新的路牌方案对 B 模型是否有准确率的提升。可以看到，尽管使用 A 模型设计路牌后，对 A 模型自己的准确率提升一般来说是最大的，但是对其他模型的识别准确率，也有进一步的提升，而且都十分显著。这里比较奇怪的是 VGG16 网络，对于自己的提升只有 4%，但是对于其他网络的提升效果却更好。这个可能是由于一定的随机性导致的，并且不影响我们的整体结论。

5.3 路牌设计思路的控制变量实验

原则 1：信息应该在纵向而不是横向

因为汽车运动时，拍摄的路牌更多会遭受到横向模糊，模糊会伤害路牌的识别准确率。如果信息更多分布在横向，即图像在 x 方向方差分布更大，就容易被模糊。所以信息应该尽可能在纵向分布，也就是交通标志在 y 方向具有更大的方差分布。因为纵向上的信息，不容易被运动模糊。

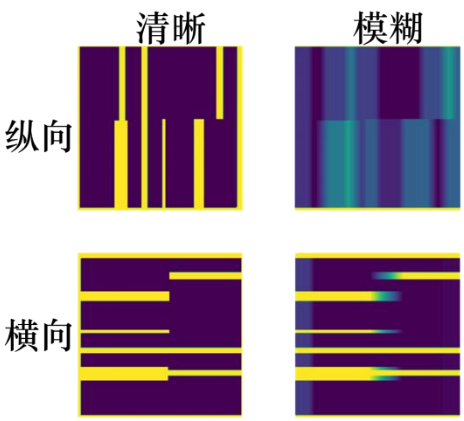


图 15 纵向和横向信息下的分类准确性

这里对结论设计了一个实验进行简单验证，使用信息都在横向方向的（纵向条纹），和信息都在纵向方向的（横向条纹）。以下验证实验均采用固定特征进行分类准确率的比较，其中采用了运动模糊对原图进行坏损操作，模糊程度随机取 50 到 300 之间的数字。可以看到模糊后横向条纹的分布还是比较清晰，而纵向条纹已经很难看清，准确率上也有显著的差距。横向条纹准确率达到 99.1%，而纵向条纹只有 58.5%。

原则 2：路牌的外轮廓应该更多样

图 16 不同轮廓丰富程度下的分类准确性

图中我们展示了实验的例子，使用三种不同情况，即形状固定、现有路牌情况，以及更丰富外形。现有路牌分类的准确率为 70.9%，当所有路牌的外轮廓都退化为正方形时准确率降为 65.7%，而使用更丰富外轮廓后，路牌识别的准确率上升到了 72.4%。

5.4 检测实验验证

在本文的大多数论述中，我们都假设路牌已经被检测模型发现和大致定位，只考虑分类的准确率。

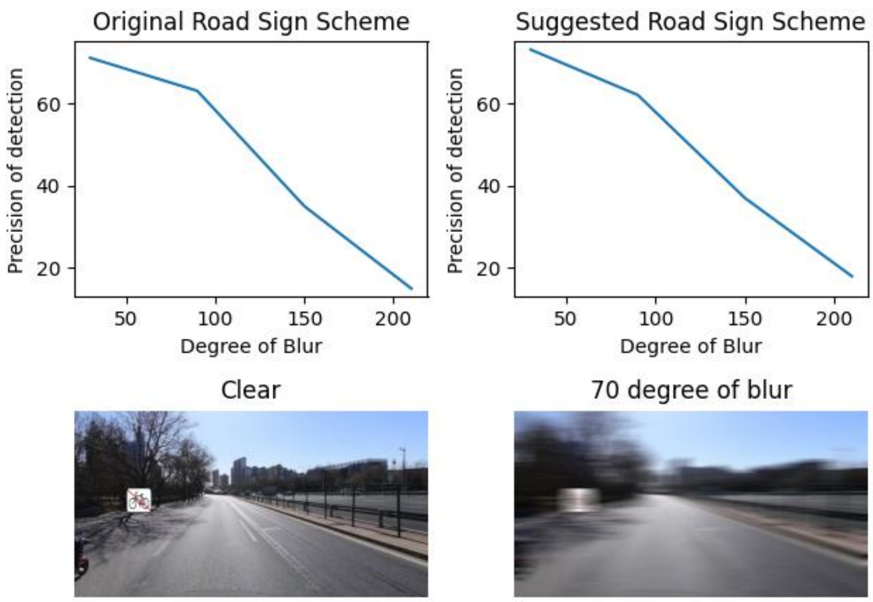


图 17 我们将设计前后的路牌”贴”到图片中进行模拟检测实验

在区域赛结束后,利用更多时间进行了补充实验。对现有路牌和本文设计方案中的路牌进行了模拟,将路牌贴到真实背景图片中,并采用基于 YOLO 的目标检测模型[5]进行检测。

实验表明,本文设计的路牌方案同样可以略微提升检测模型的性能。这可能是由于设计的路牌在多样化和视觉显著性上优于现有路牌设计所致。

[5] 我注意到，有很多同学掌握了基于 yolo 的目标检测，并且在自己的项目中进行实践。这里我也用模拟的方案，对现有的道路路牌，和本方案中设计的路牌进行了模拟，将路牌贴到真实的背景图片中，再拜托掌握检测的同学进行实验。

5.5 初步的滑轨实拍

图 18 正在进行的滑轨实拍数据采集

虽然在模拟实验上，改进的路牌取得了令人满意的效果，提升了识别准确率，但通过实际应用场景中的拍摄来验证，才更加能够令人信服。在道路真实车辆上拍摄交通路牌虽然最符合实际需求，却面临一些实际的挑战，首先很难短时间内驾车抵达每一个想要拍摄的真实路牌，更重要的是自制的路牌也不能被竖立在路边。为了复现实际应用情境，本文的实验采用了可控制的滑轨系统与小型的路牌模型。滑轨系统主要包含驱动器、电机、皮带轨道、限位器和操作台，具体结构如图 18 所示。以驱动器控制电机运行，进而带动皮带轨道旋转，从而实现对操作台移动的控制。驱动器控制运行的方向和速度，同时借助限位器，可以限制操作台运行的范围。

操作台移动的速度最高为 1.5m/s，实验拍摄中以约 0.7m/s 速度往复运行。经过等比例缩放计算，路牌约缩小了 30 倍，所以操作台移动速度对应于 21m/s 或 75.6km/h 真实车速下拍摄到路牌的状态，符合城市道路上车速较高的情况。实验拍摄中采用了木质路牌模型，并且为满足一致性，将 10 种原始路牌和算法给出的最优方案分别打印并粘贴到模型路牌上进行拍摄。

实验步骤如下：①利用现有的滑轨系统驱动操作台移动，将手机固定在操作台；②使用路牌模型，摆放在滑轨一侧，模拟出实际场景下的行驶过程；③利用手机摄像头拍摄视频，启动操作台进行移动中的拍摄；④拍摄结果中抽取出一些帧，利用开源模型进行路牌检测截取出路牌图像；⑤最终统计实验中识别准确率以及得出实验结论。结果表明原始路牌识别准确率为 78.6%，而算法设计路牌的识别准确率为 85.7%，进一步说明重新设计的路牌更容易被卷积神经网络识别。

为了便于展示，本文还利用 surface 和纸盒设计了一个简易头显，结构如图 4-c 所示，并利用代码模拟运动模糊状态。具体实现方法为获取当前图像以及缩小不同倍数后的一系列图像，并截取出公共部分进行叠加，这样叠加后就可以在静止情况下实时获取具有运动模糊效果的图片，通过人眼实际体验运动模糊下不同路牌分辨的难易程度。需要注意的是，这里仅给出视觉效果，但没有考虑不同距离造成的影响，也就是说模拟结果中可以看到运动模糊的效果，但看到的图像并不能对应一个特定速度下的观测结果。

6 结论

本文提出了通过重新设计路牌，附加易识别、难以模糊的标记来实现困难情况下交通标志识别准确率的提升。与发射信号的电子交通标志[9]相比，这种重新设计的路牌有如下优点。首先，电子交通标志的可靠性较低，因为存在潜在断电和故障风险，此时自动驾驶系统无法获得有效信息。其次，交通标志的数字化使其更容易受到网络攻击，当电子路牌的示意和真实路牌不一致时，人类无法直观地发现这个错误，使得问题更加严重。除了提升识别准确率[10,11]，防范攻击也是数字化路牌设计的一个重要的问题。本文虽然未对攻击防范展开讨论，但新设计路牌采用了更清晰和易于识别的

标识、并同时采用了更容易区分的色彩和外形轮廓，这也令不同路牌可以从多个方面进行区分，使得对路牌外观的攻击变得更加困难。

综上所述，本文总结了一些易识别路牌设计的原则，设计了新的路牌方案，并利用模拟退火算法得到了一组最优的路牌设计方案组合。同时，本文还利用实验验证了新的路牌组合比现有路牌在运动模糊情况下会有更好的识别准确率。本研究对辅助驾驶的安全性提升有一定的实践意义，特别是随着自动驾驶技术的普及，这项工作将有助于减少自动驾驶中因路牌识别错误引起的交通事故的发生。

同时，本文实际上提出了一类新的问题，即针对一个固定的分类器/分类网络，能不能在一个应用中改变类别的设计，使得分类变得更容易被识别。本文提出使用模拟退火，给出了初步的解决方案，在路牌识别领域得到了应用，并且拓展到了动漫头像设计这个问题中。这样一个能设计更有区分度的类别的算法，或许在设计领域还有着更广泛的应用。在之后的研究中，或许可以进一步把模拟退火算法进一步升级为一个端到端的算法，获得更好的效果。同时，本文后续的补充实验也验证了本算法设计出结果的跨模型有效性，那么这个设计结果是否对人眼也是有效的，也是一个值得研究的方向。

参考文献

[1] Toyota, road sign assist [EB/OL], (20230105). https://www.toyota.com.au/toyota-safety-sense/road-sign-assist .

[2] BMW,speed limit information, [EB/OL], (20230105) . http://content.bmwusa.com/microsite/3seriesActive Hybrid_2013/com/en/newvehicles/3series/sedan_active_hybrid/2011/showroom/safety/traffic_sign_recognition.html.

[3] AudiTraffic Sign Recognition[EB/OL], (20230105). https://www.audi.ca/ca/web/en/search-terms/Traffic-Sign-Recognition.html.

[4] Sermanet P , Lecun Y . Traffic sign recognition with multi-scale Convolutional Networks[C]// International Joint Conference on Neural Networks. IEEE, 2011:2809-2813.

[5] Zhe Z , Liang D , Zhang S , et al. Traffic-Sign Detection and Classification in the Wild[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[6] Wali S B , Abdullah M A , Hannan M A , et al. Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges[J]. Sensors, 2019, 19(9):2093.

[7] Kupyn O , Budzan V , Mykhailych M , et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[8] Zhu M , Han K , Zhang C , et al. Low-resolution Visual Recognition via Deep Feature Distillation[C]// ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

[9] Toh C K , Cano J C , Fernandez-Laguia C , et al. Wireless Digital Traffic Signs of the Future[J]. Iet Networks, 2018.

[10] Yuan Y , Xiong Z , Wang Q . An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition[J]. IEEE Tr ansactions on Intelligent Transportation Systems, 2017.

[11] Zhang J , Xie Z , Sun J , et al. A Cascaded R-CNN With Multiscale Attention and Imbalanced Samples for Traffic Sign Detection[J]. IEEE Access, 2020, PP(99):1-1.

附录: 动漫的实验

需要注意的是，本文的算法不一定只在自动驾驶中使用。本文实际上是提出了一种新的问题，即给定一个候选集，能不能从这里面找出一个固定大小的子集，这个子集的分类准确率是尽可能高的。这里我们用 StyleGAN 生成的动漫人物图像来进行实验。

随机选取的结果	本项目算法选择的结果

图 19 随机和算法选择的头像

从 55*3 个头像中，随机选取 55 个头像（左）与使用本文提出的模拟退火算法（右）选择的图像。在可视化时，每次都选取特征最接近的。所以可以看到左右的两个头像会比较接近。明显在随机选取的方案中，相似特征的头像对出现更多。所以在头像设计这样一个问题中，也验证了算法的价值。仔细观察可以发现，本算法可以获得一些相互之间重合度更小的人物头像。

致谢

我的项目从选题到实施再到修改、迭代，经历了一年左右的时间。在此过程中，得到了很多老师和专家的帮助，在选题阶段北京二中的高凯老师针对我提出的一些想法和创意进行指导，最终我们一起确定了本项目中的选题。2023 年初，我成功入选了中国科协和教育部面向高中学生的培养计划——“英才计划”，有幸进入到中国科学院自动化研究所复杂系统与控制国家重点实验室学习，我也要感谢实验室的刘希未研究员、宫小燕老师、赵红霞博士等给我在项目研究和实验的过程中提供的支持。我的项目曾在第 42 届北京市青少年科技创新大赛中获得一等奖，在评审环节很多专家（评审要求匿名）也针对的研究提出了很多改进意见。最后我还要感谢商汤科技的李诚研究员，在我项目遇到困境时多次鼓励我，让我继续坚持下去。