

# Multiple input self-organizing-map ResNet model for optimization of petroleum refinery conversion units

Jiannan Zhu<sup>1</sup>, Vladimir Mahalec<sup>2</sup>, Chen Fan<sup>1</sup>, Minglei Yang (✉)<sup>1,3</sup>, Feng Qian (✉)<sup>1,3</sup>

<sup>1</sup> Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

<sup>2</sup> Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L8, Canada

<sup>3</sup> Engineering Research Center of Process System Engineering, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

© Higher Education Press 2023

**Abstract** This work introduces a deep-learning network, i.e., multi-input self-organizing-map ResNet (MISR), for modeling refining units comprised of two reactors and a separation train. The model is comprised of self-organizing-map and the neural network parts. The self-organizing-map part maps the input data into multiple two-dimensional planes and sends them to the neural network part. In the neural network part, residual blocks enhance the convergence and accuracy, ensuring that the structure will not be overfitted easily. Development of the MISR model of hydrocracking unit also benefits from the utilization of prior knowledge of the importance of the input variables for predicting properties of the products. The results show that the proposed MISR structure predicts more accurately the product yields and properties than the previously introduced self-organizing-map convolutional neural network model, thus leading to more accurate optimization of the hydrocracker operation. Moreover, the MISR model has smoother error convergence than the previous model. Optimal operating conditions have been determined via multi-round-particle-swarm and differential evolution algorithms. Numerical experiments show that the MISR model is suitable for modeling nonlinear conversion units which are often encountered in refining and petrochemical plants.

**Keywords** hydrocracking, convolutional neural networks, self-organizing map, deep learning, data-driven optimization

## 1 Introduction

Large amounts of data collected daily in refining plants are a rich source of information about the performance of the processing units. Using that data to create models for monitoring and optimization of plant operations can lead to significant benefits. Due to the abundance of the heavier crude oil resource and the stricter product requirement [1], the performance of the hydrocracking unit is crucial to the profitability of a refinery [2]. Hydrocracking is a catalytic cracking process that operates under high temperature and pressure to convert heavy oil to more valuable products and remove sulfur and undesired impurities [3]. Changes in the feed properties require adjustment of operating conditions so that the yields and properties of the products will contribute to the optimal operation of the refinery. Hence, it is desirable to build an accurate model that enables accurate optimization of the hydrocracker operations.

There are two main types of methods for modeling the hydrocracking unit in general. One is the mechanistic model, which is also called the white-box model. The other is the data-driven model, also called the black-box model. In the white-box model, the description of the process is based on first principles, mechanistic or phenomenological model equations that are known [4,5]. Due to the complex composition of the feed and products and the difficulty of characterizing hydrocracking reactants, component lumping models have become a common way to describe hydrocracking units. Lumping methods divide the reaction system into several pseudo-components according to the molecular kinetic properties and establish the kinetic equations based upon these pseudo-components [6]. Discrete lumping [7–9], continuous lumping [10,11], structure-oriented lumping

Received June 29, 2022; accepted October 1, 2022

E-mails: mlyang@ecust.edu.cn (Yang M.),  
fqian@ecust.edu.cn (Qian F.)

[12], and single-event lumping are the main kinds of lumping methods [13,14].

A recent focus on modeling processing units employs deep-learning technologies by relying on a large amount of reliable historical data in a refinery. Most of the prior efforts have used a neural network (NN), or more specifically a feedforward NN (FNN) to make some predictions [15–18]. McGreavy et al. [19] adopted FNN in modeling the fluid catalytic cracking (FCC) unit to predict the yield distribution of main products and byproducts. Ochoa-Estopier et al. [20] developed an FNN model for the crude distillation column and utilized simulated annealing optimizer to maximize the revenue and minimize energy consumption. Yang et al. [21] integrated a lumped kinetic model with a traditional NN which has more hidden layers and exhibited better predictions. Recently, Song et al. [22] proposed a new method using a self-organizing map (SOM) and a convolutional NN (CNN) and achieved accurate predictions of the yield of the hydrocracking unit. The structure that Lecun et al. [23] developed was based on LeNet, a deep learning model that was used to recognize handwriting in the 1990s. A drawback of that structure is that the gradients may vanish as the network becomes deeper. In addition, Lecun et al. [23] model predicts only the yields of the products but does not predict product properties. If such model is employed to predict additional variables (e.g., properties) or if it includes more inputs, it may be easily overfitted.

In the last 10 years, numerous new deep learning techniques have emerged, especially in the fields of computer vision and natural language processing [24–27]. Within the scope of Industry 4.0, some of these methods have also been applied to the field of industrial research. In recent years, the CNN structure has become popular in fault detection and diagnosis [28–30]. Yuan and Tian [31] proposed a multiscale feature learning scheme based upon the discrete wavelet transform, CNN and long short-term memory network for fault detection and diagnosis; they verified it on the Tennessee Eastman process and *p*-xylene oxidation reaction process. Elhefnawy et al. [32] proposed an industrial fault classification that converts data into polygons based on Hamiltonian cycles first and then sends them to the CNN structure for training. Glaeser et al. [33] proposed a CNN model which contains four blocks to detect and classify fault conditions in industrial cold forging and achieved a detection and classification accuracy of more than 90%. Such deep-learning techniques have the potential to deliver improved data-driven models of refining and petrochemical process units.

This paper introduces a deep NN structure for modeling the hydrocracking unit based upon the recently developed deep-learning technologies, which can predict both the yields and properties of the products. The input variables are categorized via prior knowledge into separate inputs to the model. The model structure is divided into the

SOM part and the NN part. The former part maps the input data into multiple two-dimensional planes and sends them to the NN part. In the NN part, residual blocks are used to enhance convergence and accuracy and to ensure that the structure will not get overfitted easily [27]. In order to verify the effectiveness of the method, particle swarm optimization (PSO) and differential evolution (DE) algorithm have been applied to maximize profit by adjusting the operating conditions. The results showed that the proposed structure, i.e., multi-input SOM ResNet (MISR), performs better than Song et al. [22] in predicting product yields and leads to higher profits. Novel contributions of this work include a new deep-learning model, comprehensive comparisons among this new model, the SOM-CNN model and a typical NN model with respect to the accuracy of the training and testing, interpolation capability, and the effect when used to optimize the operating conditions. The results provide a basis for the selection of the most appropriate NN model structure for modeling conversion units such as hydrocracker or similar.

---

## 2 Hydrocracking process overview and data collection

Hydrocracking is one of the most important secondary processing units in the refinery, which converts heavy oil (e.g., wax oil and diesel) into lighter and cleaner middle distillations. Most often, the primary source of the feedstock is vacuum gas oil (VGO) and diesel. A two-stage hydrocracking process from a real refinery in China is shown in Fig. 1. The fresh feed, VGO, is mixed with hydrogen and enters the first reactor through a heater. The primary function of the first stage is hydrotreating, i.e., hydrogenation removal of metals, sulfur, nitrogen compounds, and part of the aromatic hydrocarbons in the feedstock, thereby producing a second stage feedstock with a few impurities. Typical operation achieves conversion of around 50% in the first stage by adding cracking catalyst. The purer second stage feedstock, the more likely it is that one can limit deactivation of the acid center of the cracking catalyst and be able to maintain the long-term operation of the equipment.

The bottom oil from the fractionation tower, mixed with lighter feedstock (e.g., diesel), is sent to the second stage. The task of the second stage is catalytic hydrocracking, which operates at high temperature (350–450 °C) and pressure (> 10 MPa) to convert the heavy fractions in the distillate from the first reactor into light fractions. Due to the high hydrogen partial pressure and the use of dual-function catalysts, the coking and deactivation rates are very low. The bottom products from the two reactors are sent to the fractionation train, first to a high-pressure separator, where the hydrogen-rich gas is separated and

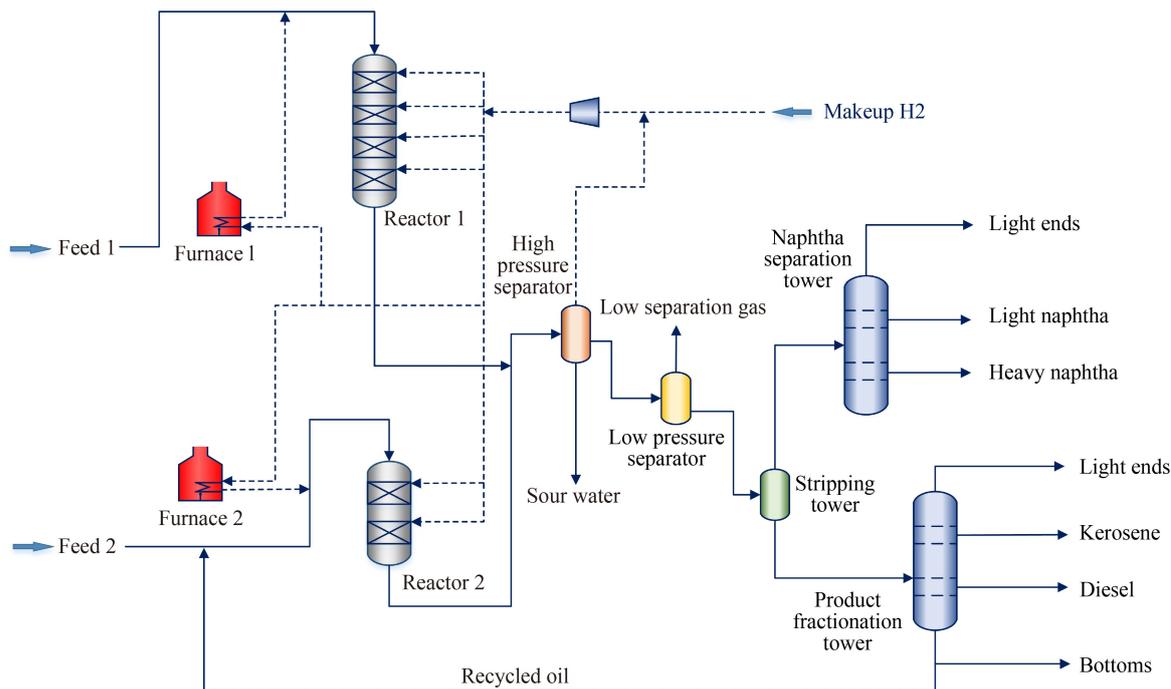


Fig. 1 A simplified flow diagram of a two-stage-series hydrocracking process.

recycled to the reactor section for mixing of fresh hydrogen feed. The separated liquid product is sent to the fractionation tower, where the gaseous product is taken out from the top of the tower, while light naphtha (LN), heavy naphtha (HN) and kerosene are removed as side products. The bottom of the fractionation tower is mixed with the hydrotreated diesel and sent to the second-stage reactor.

The complete data-driven modeling process in this work consists of five steps. The first step is data acquisition, which takes data from real refineries to determine the lower and upper limits of the input data of the HYSYS model and gathers HYSYS model outputs. The collected output data, together with the input, will be preprocessed in the second step. The second step is data preprocessing, including outlier removal, normalization to reduce the data noise, and data quality enhancement. The third step is dividing the data into a training set and a testing set that will be used to validate the accuracy of the model. The fourth step is to deliver the data into a model framework like FNN or the MISR model for training. Finally, the last step is to verify the accuracy of the various models.

### 2.1 Data generation and preprocessing

Aspen HYSYS contains models required for a rigorous simulation of the processing unit of a refinery [20]. In this work, the training and testing data are both obtained from an Aspen HYSYS model of a hydrocracking unit. Upper and lower limits of the operating conditions and feedstock properties have been obtained from a refinery in China from June 2019 to July 2021. There are several reasons

for using the simulation outputs from the Aspen HYSYS model instead of the data sets from the real refinery directly: (a) Aspen HYSYS can provide a complete set of data, while missing points may appear in the data sets collected from the database system of the refinery; (b) measuring instruments (e.g., flow and temperature) sometimes may have systemic errors; (c) stream analysis data are available infrequently and there is usually a long time interval between two analyses, usually several days or even several weeks. Moreover, the laboratory results sometimes contain errors; (d) although two years of data from the real refinery are available, around 700 daily average samples, the amount of the data may still not be sufficient to develop a deep NN model; (e) in a real refinery, the industrial data will be affected by catalyst deactivation which is difficult to model. This process can be simplified by using HYSYS models.

A total of 36 input variables and 80 output variables are considered in this work. As shown in Table 1, the input variables can be classified into three categories. The first and second categories consist of the properties of the two feeds (VGO and hydrotreated FCC diesel) separately, both containing the true boiling points (9 values), density, sulfur content and nitrogen content. The third category contains the operating conditions of the hydrocracking unit, including the ratio of the two feeds, hydrogen to oil ratio (2 values), reactor pressure, and inlet temperatures of the catalyst beds (8 values). The output variables contain the yields and properties of the eight products: sour gas ( $H_2S$ ), dry gas, LPG, LN, HN, kerosene, diesel, and tail oil (TO). The properties of the products are described in detail in the Electronic Supplementary

Material (ESM). The inputs are randomly varied between the lower and upper limits, which have been obtained from the real refinery [20]. Then, a total of 5000 data samples were generated by Aspen HYSYS. Two additional steps are needed before sending the data samples to the models for training. Since the recycled bottom flow rate is fixed to 25% of the total feed in the HCR module in HYSYS, as a consequence, when the lighter feedstock (hydrotreated FCC diesel) becomes a high proportion of the total feed, there will be a high probability that the yield of the TO will be a negative value, which is physically impossible.

Hence, the first step is that, only the samples with the TO yield within 15% to 30% have been used, yielding 3833 examples, having a distribution of yields of the products as shown in Fig. 2. In the second step, the normalization of the sample data is carried out. All input and output variables are normalized into the interval [0,1], as shown in Eq. (1). Then, scaling of the outputs of the data-driven model is as shown in Eq. (2).

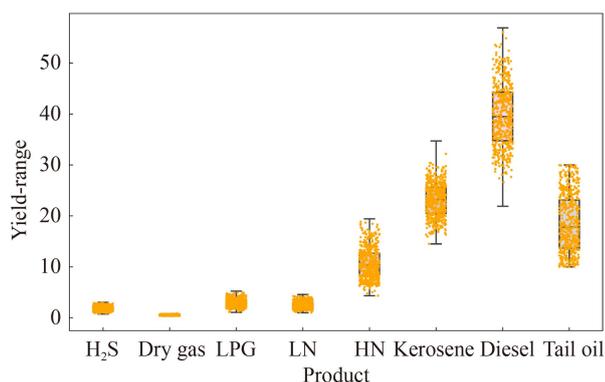
$$\widehat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

$$x = \widehat{x}(x_{\max} - x_{\min}) + x_{\min}, \quad (2)$$

where  $x$  denotes the input and output variables,  $x_{\max}$  and

**Table 1** Inputs and outputs of the model

Category	Definition	Number
Inputs-feed 1 (VGO)	True boiling points	9
	Density	1
	Sulfur content	1
	Nitrogen content	1
Inputs-feed 2 (hydrotreated FCC diesel)	True boiling points	9
	Density	1
	Sulfur content	1
	Nitrogen content	1
Inputs-operating conditions	Feed ratio	1
	Hydrogen to oil ratio	2
	Reactor pressure	1
	Inlet temperatures	8
Outputs	Yields	8
	Properties	72



**Fig. 2** Distribution of raw data (yields of the eight products).

$x_{\min}$  denote the maximum and minimum values of these variables.

Randomly selected 3000 samples have been used for training and the rest have been used for testing. That is, the testing data accounts for 21.7% of the total data. We take 3000 training data examples (integer multiples of one thousand) in order to facilitate the testing of different batch sizes when training the NN models.

### 3 MISR model

As mentioned in the Introduction, the SOM-CNN model was first proposed by Song et al. [22] and applied in the prediction of product yields and bed temperatures of the hydrocracking unit. SOM in the SOM-CNN framework maps high-dimensional data into a two-dimensional plane. SOM without training can increase the information entropy in the local area and thus improve the accuracy of the prediction. The conventional NN, usually containing convolutional, pooling and fully connected layers, is commonly used for feature extraction and classification of images and is able to achieve high accuracy. Popularity of CNN started with LeNet in 1998 to solve the task of handwritten digit recognition [23] and flourished over the last decade with the emergence of AlexNet (2012) [25], GoogLeNet (2015) [26], VGGNet (2015) [34], ResNet (2016) [27], etc. With the advancement of computer hardware capabilities, CNN networks are becoming more structurally complex and better able to implement more complex classification and prediction tasks.

In Song's work, the high-dimensional input vectors are mapped into a two-dimensional plane represented by a  $28 \times 28$ -size SOM and sent into the CNN part. The CNN structure consecutively contains two convolutional and pooling layers, followed by a fully connected layer. This structure is highly similar to LeNet proposed by LeCun [23]. Compared with the conventional FNN, the absolute prediction errors of heavy products like kerosene, diesel, and TO have been reduced by 0.1%, and the absolute mean error of the total products has been decreased to 0.36%. However, the premise of Song's work is that the data are continuous, less fluctuating, and only eight products' yields need to be predicted. When the same framework is applied to a more fluctuating and discontinuous data set to predict both the yields and properties of the products, the results are unsatisfactory. This has been confirmed by a series of experiments carried out in the present work.

This work introduces a different deep NN structure that resolves issues associated with the SOM-CNN model.

#### 3.1 Multi-input-SOM-CNN

In computer vision processing, a colorful image will be

split into red, green, and blue channels and then sent to the conventional NN. Inspired by this, prior knowledge of the processing unit can be utilized to divide the input variables into multiple groups and then map them into several different two-dimensional planes via different SOM weights, respectively.

The two parts of the proposed MISR framework can be seen in Fig. 3. As shown in Fig. 3(a), in the SOM part, the 36-dimensional inputs will be split into three sets. The first and second sets are the variables for the two different feeds, and the third one contains the operating conditions, as listed in Table 1. Then, three  $32 \times 32$  sized channels will be obtained via three training-free SOM weights and sent to the CNN structure. The selection of SOM size is determined by the execution time and the accuracy. As listed in Table 2, five different structures ranging from  $24 \times 24$  to  $128 \times 128$  have been evaluated.  $R^2$  represents the correlation coefficient of the 13 output variables. It is evident that the execution time increases exponentially, and the accuracy increases with the size of SOM. For the balance between computing time and accuracy,  $32 \times 32$

was selected as the size of the SOM. It is worth noting that all experiments in this work except the optimization part in Section 5 are based on a Linux server with a Xeon gold 6240 2.60 GHz CPU and a Nvidia 2080 TI GPU.

### 3.2 Residual blocks

The ResNet proposed by He et al. [27] won imagenet competition in 2015. As shown in Fig. 4, compared with the traditional CNN structure (left), each block of the ResNet (right) has an additional map from input to output. The core concept of ResNet is that each additional layer should contain the original function easily. Hence, the new model will be at least equivalent to or more effective than the previous model when adding new layers to the NN to reduce the training error. Many research and commercial network structures have been influenced by ResNet [35,36].

In this work, the residual blocks are used for the reconstruction of the NN structure. As shown in Fig. 3(b), each residual block contains a convolutional layer, a

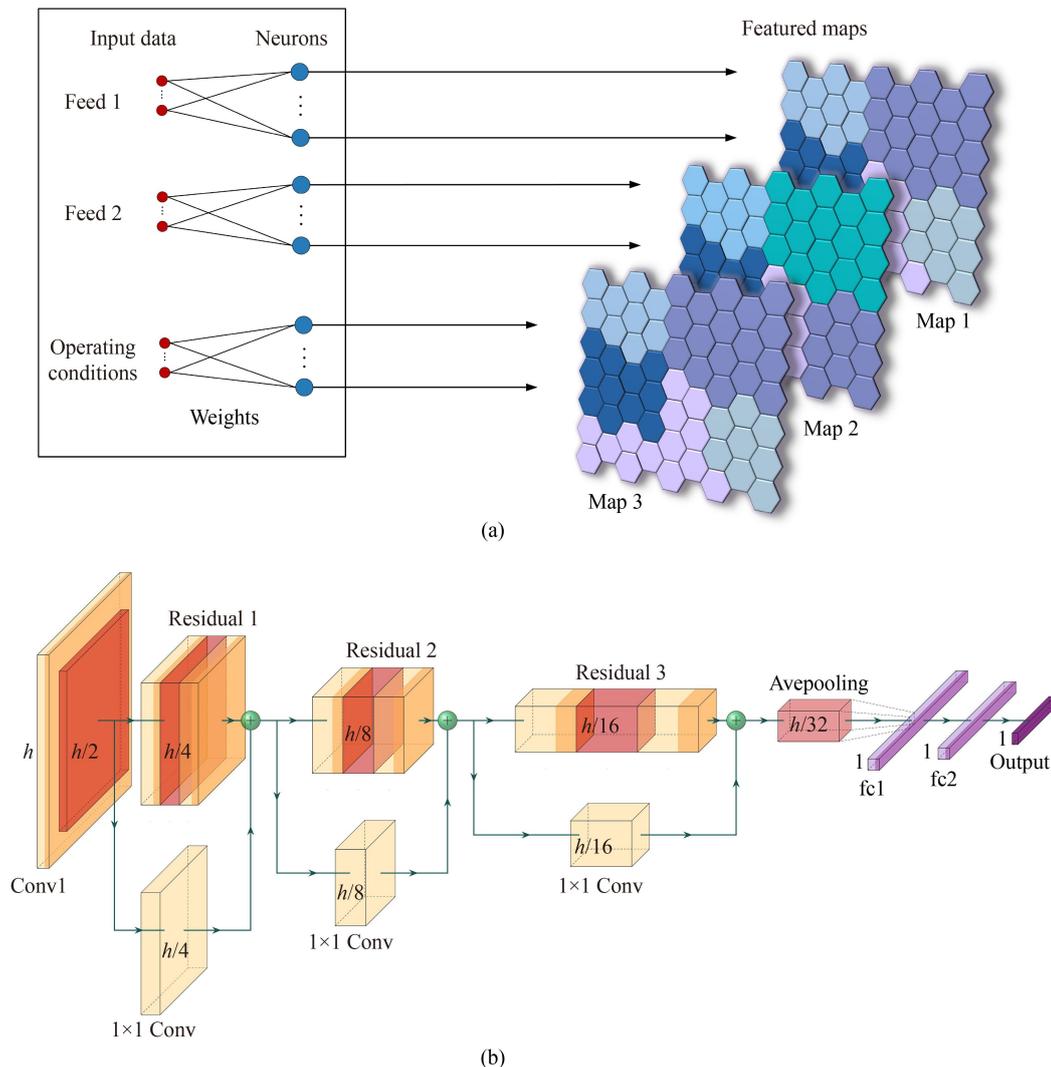


Fig. 3 Structure of MISR of (a) SOM part and (b) residual part (3 residual blocks).

batch normalization (BN) layer, a pooling layer, and a convolutional layer in sequence. The number of input and output channels of the residual blocks is the same to ensure the input can be added to the output. In this work, models containing 2–6 residual blocks have been evaluated. It has been found that the training time increases exponentially as the number of residual blocks increases, and the error barely decreases when it reaches 5 residual blocks. The model with three residual blocks is recommended in the proposed framework and employed in the following experiments for comparison.

### 3.3 MISR framework

As shown in Fig. 3, the proposed MISR framework can be divided into two parts: the SOM part and the ResNet part. Figures 3(a) and 3(b) represent the SOM part and the NN part, respectively. The SOM maps an input vector into two-dimensional planes. In this work, specifically, the three groups of input variables were mapped by three  $32 \times 32$  SOMs to obtain three  $32 \times 32$  two-dimensional data. Significantly, the weights of three SOMs were randomly initialized in the range of  $[0,1]$  and remained unchanged, which is called “train-free”. The main advan-

tage of “train-free” SOMs is increasing the information entropy to improve the final prediction accuracy [22]. It should be mentioned that the outputs of the SOMs are expected to be approximately  $n/2$  times the input. Thus, it needs to be multiplied by  $2/n$  before being sent to the ResNet part, in which  $n$  represents the number of variables in an input vector.

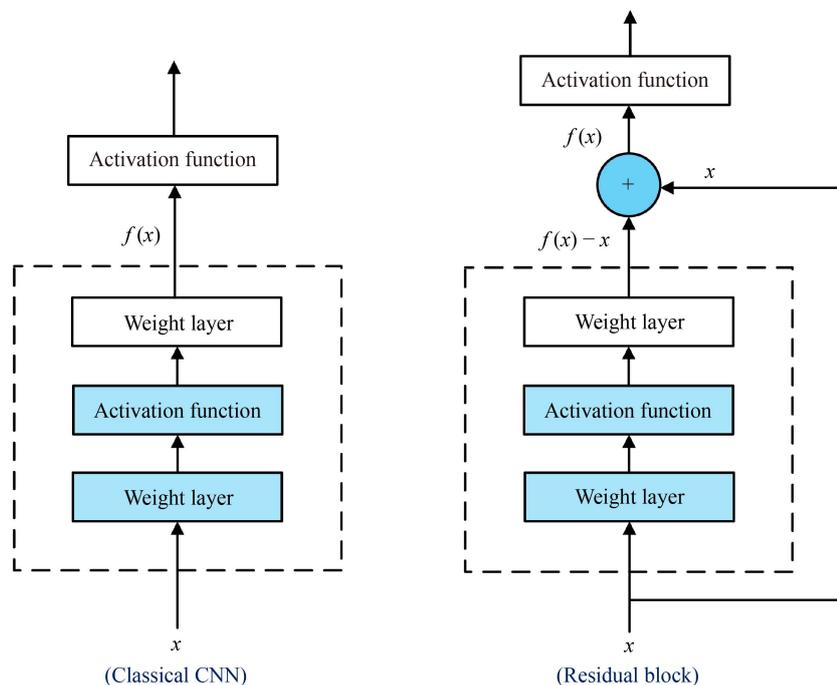
The ResNet part stacks the residual blocks. As is shown in Fig. 3(b), the architecture of the adopted ResNet part is conv-pooling-residual blocks-pooling-fc-fc-output. In a single residual block, a  $3 \times 3$  kernel will half the resolution (which is the size of the two-dimensional input data) in a stride of 2, followed by a BN layer. ReLU was chosen as the activation function applied immediately behind the BN layer, followed by another convolutional layer. The additional convolutional layer was used to identity mapping the input data. The output of the former convolutional and the additional layer will be added and sent to the activation function.

## 4 Training and comparison of SOM-CNN, MISR, and FNN models

The following sections will analyze the impact of network structure on model accuracy and convergence. In addition, the experiments listed below predict only 13 key outputs, including the 8 flow rates of the product, the octane number of the LN and HN, the smoke point of kerosene, and the cetane number of diesel, and the bureau of mines correlation index of the TO. Comparative results for predicting all 80 output variables are presented in the ESM.

**Table 2** Statistics related to performances of different SOM sizes under 2000 iterations (prediction of 13 outputs)

Index	24 × 24	32 × 32	48 × 48	96 × 96	128 × 128
Executing time/min	7.58	8.09	8.51	17.33	32.99
$R^2$ (correlation coefficient)	0.950	0.956	0.953	0.9611	0.9606
Mean relative error (MRE)	2.3587	1.9645	2.2186	1.8196	1.6652
Mean absolute error (MAE)	0.4263	0.3610	0.3768	0.3171	0.2960



**Fig. 4** Residual block (right) compared with the classical CNN (left).

#### 4.1 Influence of BN

BN is a method used to make artificial NNs faster and more stable through the normalization of the layers' inputs by re-centering and re-scaling. It was first proposed by Ioffe in 2015 [24] and then widely used in various efficient NN structures. This section describes experiments comparing the original SOM-CNN structure and the SOM-CNN structure with the BN layers.

The SOM in two contrastive experiments will be both free of training for increasing the information entropy to enhance the accuracy of the prediction. The CNN part of the former structure is a bit different from the one proposed by Song et al. [22], the active function used is ReLU, and the pooling method is average pooling. The BN layers are added behind each active function. Ten independent runs have been performed for the two frameworks, and the results have been averaged. There are 2000 iterations in each run.

The convergence curves of the two structures can be seen in Fig. 5. It is clear that the loss function of SOM-CNN with BN converges quickly at the beginning of the iterations. However, SOM-CNN without BN converges slowly in the first 750 generations, with sharp fluctuations occurring in the process. In addition, the ultimate loss of the SOM-CNN with BN is a bit lower than the one without BN. As shown in Table 3, although the training time per round of the structure with BN is two minutes longer than the one without BN, the accuracy of the former is better than the latter. The correlation coefficient of the former can reach 0.9468, which is 0.014 higher than the latter. The mean relative and absolute errors of the 13 outputs of the structure with BN are reduced by 8.9% and 10.2% compared with those without BN, reaching 2.316 and 0.4416, respectively.

As a commonly used method in modern NNs, the utilization of BN can indeed improve convergence and accuracy, especially for the data set with a large fluctuation range as in the hydrocracker data set used in this work. This result has motivated the use of BN in the subsequent residual blocks and fully connected layers to improve the

convergence rate and accuracy of the overall MISR network.

#### 4.2 Influence of multi-input channels

In this section, the effect of using multi-input is examined. One of the two contrastive models tested in this section was the same as the SOM-CNN model with BN mentioned in the previous section. In the other model, 36 input variables were classified into three groups and then sent to the three SOMs mentioned in Section 3.1, the other structure remaining the same. Ten independent runs were performed for two frameworks, with 2000 iterations for each run.

The convergence curves of the two structures can be seen in Fig. 6. The loss of the two frameworks are not much different because BN layers are added to both two CNN structures. However, the convergence speed of the structure with multi-input is still a bit faster than the other, and the ultimate loss is a bit lower in addition. As can be seen in Table 4, although the training time per round is 0.35 min longer, the accuracy of the framework with multi-input is better than the one without. The correlation coefficient of the former is 0.9533, which is 0.69% higher than the latter. The mean relative and absolute errors of the 13 outputs of the structure with multi-input are reduced by 8.59% and 13.3% compared with the one without multi-input, reaching 2.117 and 0.3870, respectively.

It can be concluded that classifying the input variables by prior knowledge in advance and mapping them via the SOMs to obtain the input channels of the NN structure is an effective modeling method. This is likely due to the reduction of model complexity caused by separating the sets of input variables based on prior knowledge.

#### 4.3 Comparison of MISR, FNN, and SOM-CNN frameworks

In this section, residual blocks will be introduced to modify the original CNN structure to enhance the

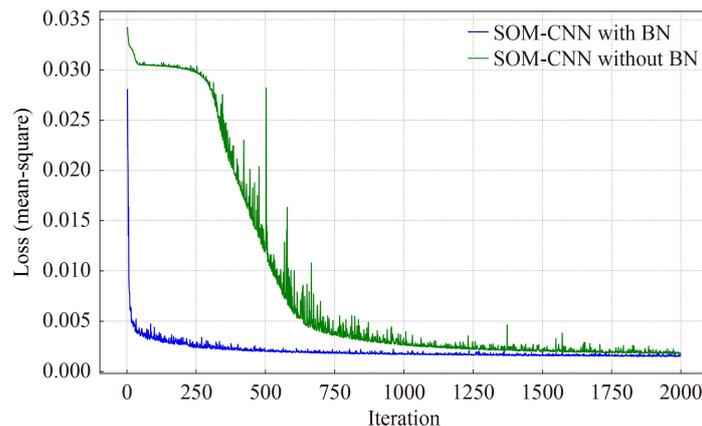


Fig. 5 The loss of SOM-CNN with and without BN.

stability and robustness of the framework. The resulting network MISR is compared with the classical FNN and the SOM-CNN structure to analyze the convergence and accuracy of this proposed structure.

The MISR structure applied contains three layers of the residual blocks mentioned in Section 3.3. The comparison of the real data and the prediction data of MISR for the 8 product yields of the hydrocracking unit on a test set of 883 samples is shown in Fig. 7, where the  $x$ -axis represents the predicted value of the MISR structure and the  $y$ -axis represents the actual value simulated by HYSYS. It can be seen that the scatter points follow the  $y = x$  line closely, implying the prediction is accurate enough. In addition, the correlation coefficients of the prediction for each product can be seen in this figure. The correlation coefficients of most of the predictions can reach around 0.98. Absolute error bands are added to each prediction to visualize the errors clearly. In addition, the same structure with more residual blocks has been tested (see Table 5). It can be seen that the error gradually decreases as the number of residual blocks increases until reaching 5 residual blocks. The highest accuracy that the structure can achieve has 0.9655 of the average correlation coefficient, and the mean relative and absolute errors of 1.6740 and 0.3071, respectively.

The classical FNN models with 1 to 5 hidden layers and different structures have been tested to evaluate their competitiveness with the proposed framework. As shown in Table 6, the parameters of the NN grow rapidly and the accuracy of the model improves as the number of hidden layers increases. However, when the number of hidden layers was increased from 4 to 5, the number of network parameters was doubled, and there was little improve-

ment in accuracy, with the MRE reduced by only 0.001.

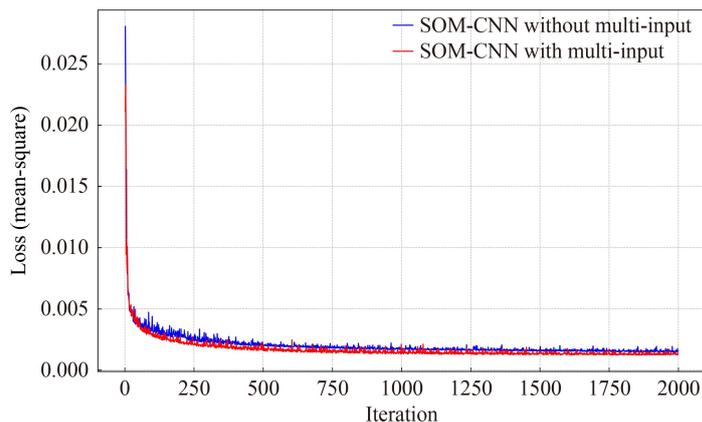
The classical FNN model with 3 hidden layers was chosen as the “baseline” to analyze the improvement that MISR made compared with the SOM-CNN and multi-input-SOM-CNN frameworks. The structure of the tested SOM-CNN follows closely the structure in Song’s work [22], except that the size of the SOM was expanded to  $36 \times 36$  to accept more input variables. The structure of the multi-input-SOM-CNN is described in Section 4.2. Data presented in Table 7 show that the prediction results of the classical FNN and SOM-CNN are similar in terms of the correlation coefficient, mean absolute and relative errors. The multi-input-SOM-CNN can improve a little bit on the SOM-CNN in terms of the same three indicators. What stands out in this table is that compared with the original SOM-CNN, the MISR framework has a 2.3% improvement in the correlation coefficient, and the relative error decreased from 2.33% to 1.68%. The absolute error decreased from 0.46 to 0.31.

The test error curves of the four frameworks mentioned above are shown in Fig. 8. It is apparent from the figure that the SOM-CNN structure has a more dramatic fluctuation during the convergence process. On the other hand, the convergence of multi-input-SOM-CNN is much smoother, mainly due to the addition of the multi-input structure and BN. What stands out in this figure is that the MISR structure with only three residual blocks has faster convergence, less fluctuation, and lower final loss.

The interpolation test of the MISR model for different feed ratios has been carried out to verify the interpolation ability of the model and the ability to judge the trend of product yields when operating conditions change. This information is also useful when using the model for the optimization of operating conditions (in Section 5). The results of the comparison with the HYSYS and the classical FNN models can be seen in Fig. 9. Though MISR has more slight fluctuations, it captures the trend when the feed ratio changes. In particular, when predicting the yield of the HN, the FNN model gave the opposite trend to HYSYS, while the trend given by MISR was consistent.

**Table 3** Comparison of SOM-CNN with and without BN

Index	SOM-CNN without BN	SOM-CNN with BN
Iterations	2000	2000
Correlation coefficient $R^2$	0.9329	0.9468
MRE (10 samples)	2.543	2.316
MAE (10 samples)	0.4970	0.4464
Time cost/min	5.50	7.66



**Fig. 6** The loss of SOM-CNN with and without multi-input.

These results suggest that the MISR framework inherits the superimposed effect of the advantages of the BN

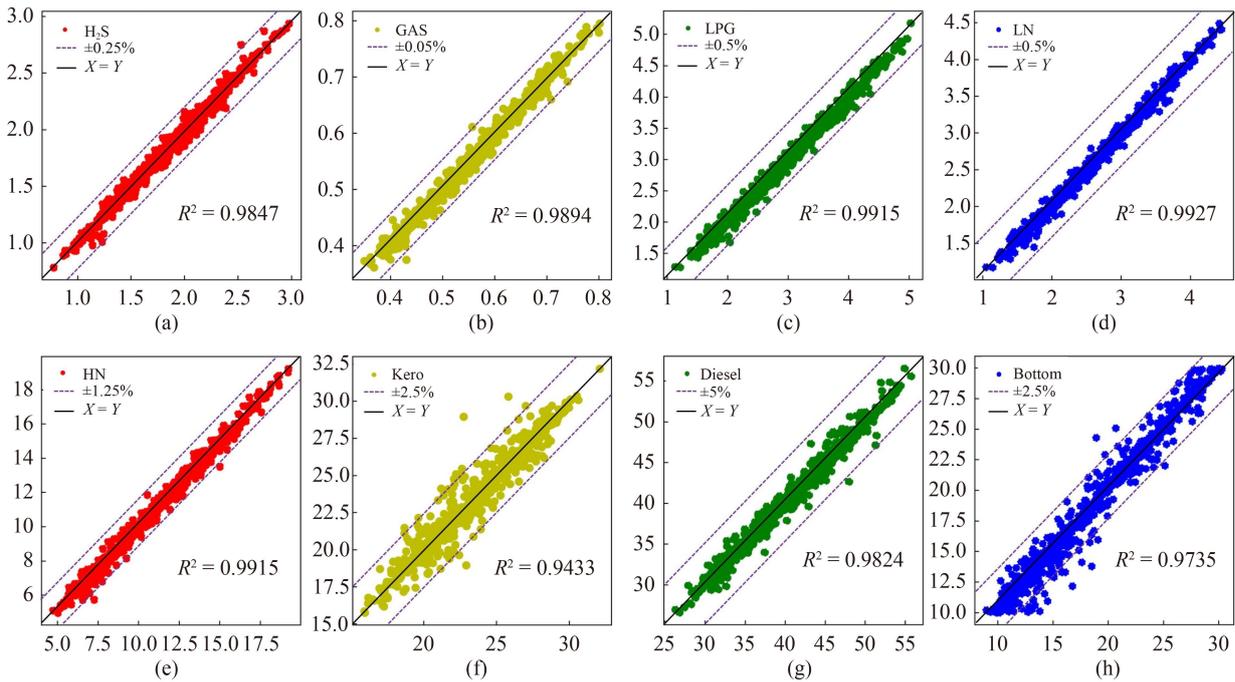
technology, the multi-input structure, and the residual blocks, having satisfactory fitting accuracy and convergence characteristics. It is likely that MISR is capable of modeling complex refinery processes.

**Table 4** Comparison of SOM-CNN with and without multi-input

Index	SOM-CNN	Multi-input-SOM-CNN
Correlation coefficient $R^2$	0.9468	0.9533
MRE (test samples)	2.316	2.117
MAE (test samples)	0.4464	0.3870
Time cost/min	7.66	8.01

## 5 Optimization of hydrocracker operation

In order to verify the applicability of the MISR



**Fig. 7** The predicted yields vs. actual yields of the eight products on the testing data: (a) H<sub>2</sub>S, (b) GAS, (c) LPG, (d) LN, (e) HN, (f) kerosene, (g) diesel, and (h) bottom.

**Table 5** Performances of MISR with multiple residual blocks

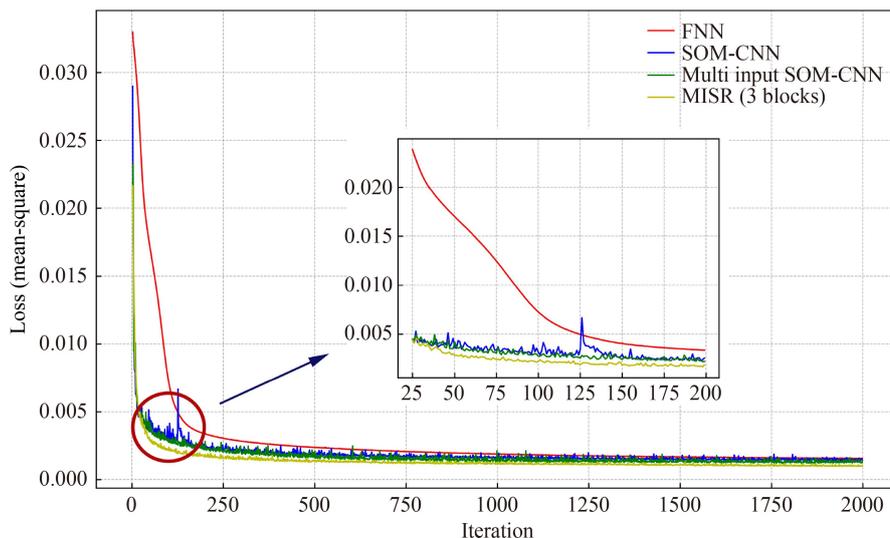
Index	2 residual blocks	3 residual blocks	4 residual blocks	5 residual blocks
Loss	0.00104	0.00102	0.00102	0.00100
Iterations	2000	2000	2000	2000
Total time/min	19.7	25.9	33.1	70.8
Correlation coefficient $R^2$ (total outputs)	0.9628	0.9635	0.9638	0.9655
$R^2$ (properties only)	0.9369	0.9354	0.9368	0.9402
MRE (test samples)	1.862	1.6928	1.6710	1.6700
MAE (test samples)	0.3371	0.3177	0.3195	0.3071
Number of trainable parameters	378,413	1,580,000	6,379,000	25,563,000

**Table 6** Performances of classical FNN models with different hidden layers

Index	1 hidden layer	2 hidden layers	3 hidden layers	4 hidden layers	5 hidden layers
Structure	36-64-13	36-128-64-13	36-128-128-64-13	36-128-256-128-64-13	36-128-256-256-128-64-13
Loss	0.00211	0.00174	0.00161	0.00125	0.00120
Iterations	5000	5000	5000	5000	5000
Total time/min	3.8	5.2	6.2	9.4	10.3
Correlation coefficient $R^2$ (total outputs)	0.928	0.939	0.943	0.954	0.956
$R^2$ (properties only)	0.868	0.890	0.895	0.915	0.917
MRE (test samples)	2.623	2.328	2.249	1.999	1.914
MAE (test samples)	0.518	0.470	0.455	0.4187	0.4035
Number of trainable parameters	3213	13837	30349	96269	145549

**Table 7** Performances of different networks

Index	FNN	SOM-CNN	Multi-input-SOM-CNN	MISR with 3 residual blocks
Loss	0.00161	0.00157	0.00127	0.00099
Iterations	5000	2000	2000	2000
Total time/min	6.2	7.6	8.1	25.2
Correlation coefficient $R^2$ (total outputs)	0.9434	0.9418	0.9536	0.9638
$R^2$ (properties only)	0.895	0.892	0.917	0.937
MRE	2.249	2.332	1.987	1.686
MAE	0.455	0.456	0.386	0.314

**Fig. 8** Loss curves of the four networks.

hydrocracker model, the operation of the hydrocracker unit has been optimized via PSO and DE optimization algorithms. The objective function, which contains both the yields and properties of the products, is calculated by

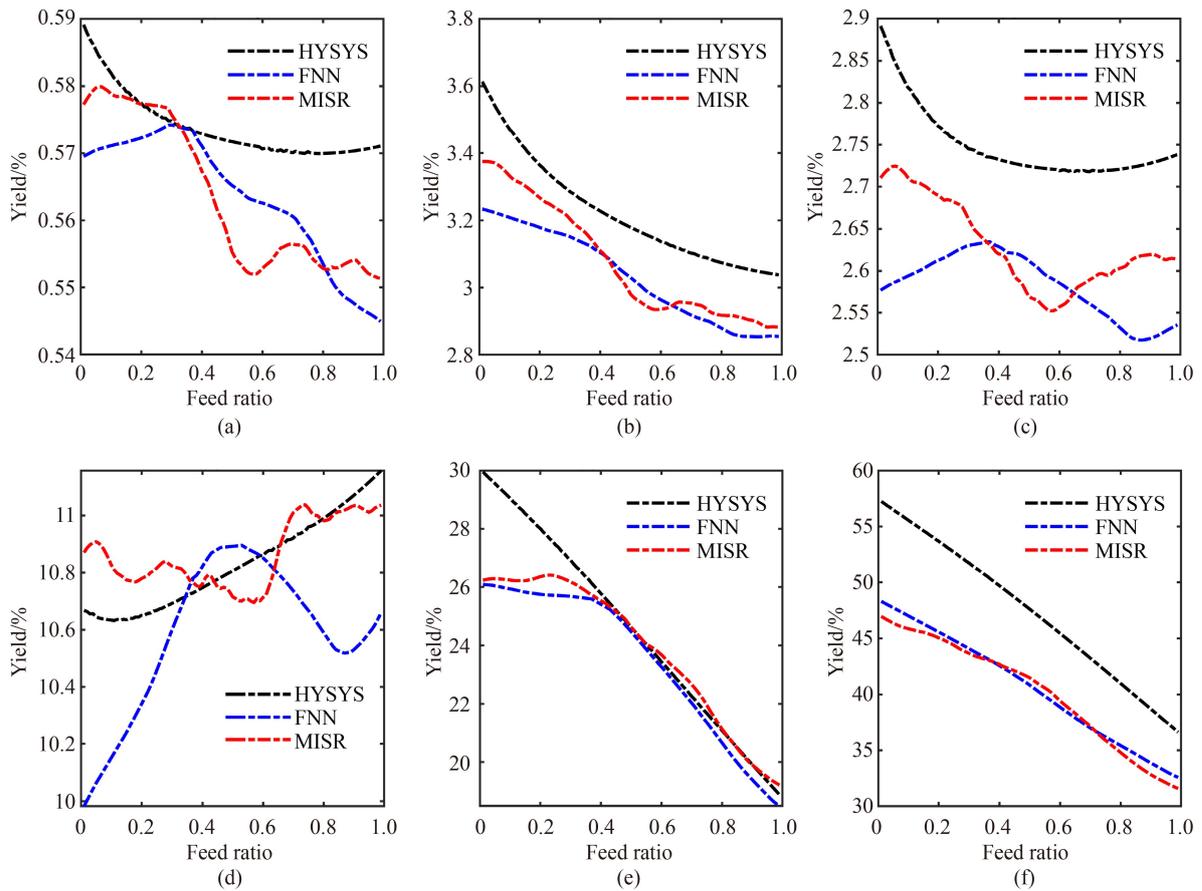
$$\begin{aligned}
 Q = & 462 \times SG + 2200 \times G + 2200 \times LPG + (1378 \\
 & + (RON_{LN} - 7500)) \times LN + (1168 + (RON_{HN} \\
 & - 5660)) \times HN + (1442 + (19.8 - 1000 \times SP)) \\
 & \times KE + (2962 + 100 \times (Ce - 40)) \times DI \\
 & + (1075 + 200 \times (30 - BMCI)) \times TO, \quad (3)
 \end{aligned}$$

where  $SG$ ,  $G$ ,  $LPG$ ,  $LN$ ,  $HN$ ,  $KE$ ,  $DI$ , and  $TO$  denote the product yields of sour gas, dry gas, LPG, LN, HN, kerosene, diesel, and TO, respectively. The coefficients of all yields are the prices of these products. Each product price has been calculated in two parts: one is the original price obtained from an average market price for a specific time period, and the other is the price penalty factor given by the key properties of the products obtained by estimation. For each product, we selected its key property to calculate the price penalty factor. In the formula,  $RON_{LN}$  and  $RON_{HN}$  denote the research octane number (RON) of LN and HN.  $SP$  denotes the smoking point of kerosene, and  $Ce$  denotes the cetane number of diesel.  $BMCI$  is the bureau of mines correlation index of the TO. An example of how to calculate the price penalty factor of gasoline  $RON_{LN}$  is shown in Eq. (4):

$$RON_{LN} = \frac{P_{95\#} - P_{92\#}}{O_{95\#} - O_{92\#}}, \quad (4)$$

where,  $P_{95\#}$  and  $P_{92\#}$  denote the price of 95# and 92# gasoline.  $O_{95\#}$  and  $O_{92\#}$  denote the octane number of 95# and 92# gasoline, respectively. That is,  $RON_{LN}$  equals the price deviation of two gasoline types (95# and 92# gasoline) divided by the octane number deviation. All of these products and properties have been calculated by the MISR and SOM-CNN models for comparison purposes. The operating conditions are optimized within the specific range, consistent with the range of the input variables mentioned in Section 2.1. It should be noted that the operating cost and utility cost were ignored in the following experimental cases.

Three implementations of optimization algorithms have been compared: (i) single DE, (ii) single PSO, and (iii) 40-round-PSO. The last one was chosen as the final optimization algorithm since it exhibited acceptable execution times and stable optimal results, as is shown in Table 8. Compared with PSO, DE has better population diversity and can find better solutions in a single round. However, the disadvantage is a longer execution time, up to 9.25 s per iteration, due to the special selection-crossover-mutation process. PSO is simpler and has a much faster convergence speed, taking only 0.31 s per iteration step. However, it converges to a local optimum



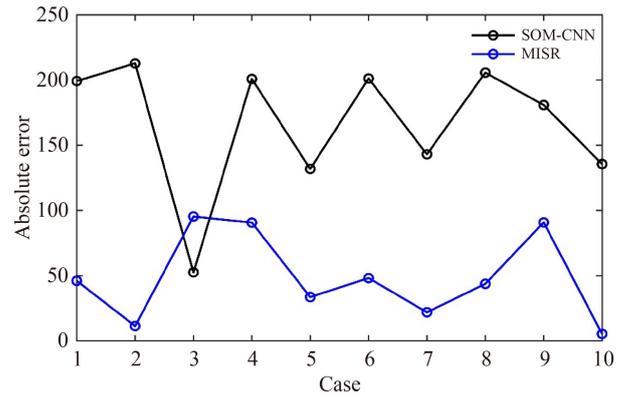
**Fig. 9** Interpolation test of different models to predict yields of the six products based on different feed ratios: (a) dry gas, (b) LPG, (c) LN, (d) HN, (e) kerosene, and (f) diesel.

**Table 8** Optimization effects of three methods

Index	1 round-PSO	1 round-DE	Multi-round-PSO
Rounds	1	1	40
Iterations per round	400	1000	400
Total time/min	2.06	154.16	82.54
Seconds per iteration	0.31	9.25	0.31
Max profit (10 times)	2420.39	2413.02	2429.67
Mean profit (10 times)	2321.91	2395.19	2423.19

easily, having a worse average optimal result. Multi-round-PSO takes advantage of the fast convergence of PSO, while improving population diversity by multiple rounds. The result has proven to be better than DE, and the execution times are about half as long as the single-round DE method. Consequently, 40-round-PSO has been chosen as the main method, while DE has been used to verify the results. All calculations in this section have been carried out on i7-8700 CPU 3.2 GHz and 16 G memory.

Ten real cases which differ in the feed properties were chosen to be optimized. In each case, the properties of the two feeds (TBP, density, sulfur, and nitrogen) were fixed. The operating conditions, including the feed ratio, two hydrogen oil ratio, reactor pressure, and eight reactor bed temperature, were optimized. Optimal operation



**Fig. 10** Difference between true optimum profit and profit by SOM-CNN and MISR.

conditions were entered into the HYSYS model to obtain the ‘real’ products information, and then the ‘real’ profit was calculated. As is shown in Fig. 10 and Table 9, MISR locates better profit than SOM-CNN. The prediction error of MISR (48.28 on average) is nearly 1/3 of SOM-CNN (158.36 on average). MISR identifies higher profit in cases 1, 3, 4, 6, 7, 9, 10, resulting in higher average profit (2228.98) than SOM-CNN (2211.38). The results show that the profit based on the MISR model is significantly

**Table 9** Profit prediction and real optimization results via SOM-CNN and MISR

Case number	SOM-CNN prediction benefit	SOM-CNN real benefit	SOM-CNN prediction error	MISR prediction benefit	MISR real benefit	MISR prediction error
Case 1	2068.98	2268.22	199.24	2232.38	2278.4	46.02
Case 2	2066.36	2279.12	212.76	2224.89	2236.27	11.38
Case 3	2036.59	2089.14	52.55	2238.26	2142.88	−95.38
Case 4	2076.25	2277.07	200.82	2207.9	2298.58	90.68
Case 5	2040.76	2172.62	131.86	2192.5	2158.91	−33.59
Case 6	2065.55	2186.73	121.18	2243.26	2291.41	48.15
Case 7	2034.07	2177.14	143.07	2165.76	2187.69	21.93
Case 8	2041.77	2247.44	205.67	2200.91	2244.65	43.74
Case 9	2048.19	2229.03	180.84	2155.81	2246.63	90.82
Case 10	2051.72	2187.32	135.6	2203.24	2204.42	1.18
Mean	2053.024	2211.383	158.36	2206.491	2228.984	48.28

more accurate than SOM-CNN due to the fact that it can predict better both yields and properties.

## 6 Conclusions

This study proposes a new deep learning model, i.e., MISR, for modeling hydrocracking units in petroleum refineries. Compared with the SOM-CNN structure, this model is distinguished by the use of prior knowledge to differentiate between different types of network input data and by the introduction of BN technology and residual blocks. These changes to the model structure enable better accuracy, convergence and robustness of the model. Multiple comparative experiments confirm the significance of the multi-input method, the BN technology, and the utilization of the residual blocks. Compared with the SOM-CNN structure, the MISR framework has a 2.3% improvement in the  $R^2$  (correlation coefficient), the MRE decreases from 2.33% to 1.68%, and the MAE decreases from 0.46 to 0.31. In addition, MISR exhibits a more stable decreasing trend of prediction error during the training process. The MISR model of a hydrocracking unit has been optimized via multi-round-PSO and DE algorithms. The results show that the prediction of the optimal operation based upon the MISR structure is much more accurate, and the objective function is higher than those computed via the previously published SOM-CNN model.

Results of this work suggest that the MISR model is an accurate and robust data-driven model with excellent convergence. It is shown to have a better performance in predicting the complicated properties (more than 70) of products than the typical FNN model and the SOM-CNN model. Provided an adequate amount of industrial data can be used for training, the MISR model will be easier to build and more accurate. In contrast, the HYSYS model is not easy to build and its kinetic parameters need to be constantly adjusted, which is a cumbersome trial and error procedure. Besides, it is likely that the MISR framework can accurately model other types of

conversion units in the refining and petrochemical plants.

**Acknowledgements** This work was supported by the National Natural Science Fund for Distinguished Young Scholars (Grant No. 61725301), the National Natural Science Foundation of China (Basic Science Center Program: Grant No. 61988101), International (Regional) Cooperation and Exchange Project (Grant No. 61720106008), and General Program (Grant No. 61873093).

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at <https://dx.doi.org/10.1007/s11705-022-2269-5> and is accessible for authorized users.

## References

- Marafi A, Albazzaz H, Rana M S. Hydroprocessing of heavy residual oil: opportunities and challenges. *Catalysis Today*, 2019, 329: 125–134
- Iplik E, Aslanidou I, Kyprianidis K. Hydrocracking: a perspective towards digitalization. *Sustainability*, 2020, 12(17): 7058
- Ward J W. Hydrocracking processes and catalysts. *Fuel Processing Technology*, 1993, 35(1): 55–85
- Sánchez S, Rodríguez M A, Ancheyta J. Kinetic model for moderate hydrocracking of heavy oils. *Industrial & Engineering Chemistry Research*, 2005, 44(25): 9409–9413
- Kumar H, Froment G F. Mechanistic kinetic modeling of the hydrocracking of complex feedstocks, such as vacuum gas oils. *Industrial & Engineering Chemistry Research*, 2007, 46(18): 5881–5897
- Félix G, Ancheyta J. Using separate kinetic models to predict liquid, gas, and coke yields in heavy oil hydrocracking. *Industrial & Engineering Chemistry Research*, 2019, 58(19): 7973–7979
- Singh J, Kumar M, Saxena A K, Kumar S. Reaction pathways and product yields in mild thermal cracking of vacuum residues: a multi-lump kinetic model. *Chemical Engineering Journal*, 2005, 108(3): 239–248
- Qader S, Hill G. Hydrocracking of gas oil. *Industrial & Engineering Chemistry Process Design and Development*, 1969, 8(1): 98–105
- Bhutani N, Ray A K, Rangaiah G. Modeling, simulation, and multi-objective optimization of an industrial hydrocracking unit. *Industrial & Engineering Chemistry Research*, 2006, 45(4):

- 1354–1372
10. Laxminarasimhan C S, Verma R P, Ramachandran P A. Continuous lumping model for simulation of hydrocracking. *AIChE Journal*, 1996, 42(9): 2645–2653
  11. Lababidi H M S, AlHumaidan F S. Modeling the hydrocracking kinetics of atmospheric residue in hydrotreating processes by the continuous lumping approach. *Energy & Fuels*, 2011, 25(5): 1939–1949
  12. Quann R J, Jaffe S B. Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures. *Industrial & Engineering Chemistry Research*, 1992, 31(11): 2483–2497
  13. Becker P J, Serrand N, Celse B, Guillaume D, Dulot H. Comparing hydrocracking models: continuous lumping vs. single events. *Fuel*, 2016, 165: 306–315
  14. Becker P J, Serrand N, Celse B, Guillaume D, Dulot H. A single events microkinetic model for hydrocracking of vacuum gas oil. *Computers & Chemical Engineering*, 2017, 98: 70–79
  15. Rosli M, Aziz N. Review of neural network modelling of cracking process. In: *Second International Conference on Chemical Engineering (ICCE)*. Bandung, Indonesia: IOP, 2016
  16. Bhutani N, Rangaiah G P, Ray A K. First-principles, data-based, and hybrid modeling and optimization of an industrial hydrocracking unit. *Industrial & Engineering Chemistry Research*, 2006, 45(23): 7807–7816
  17. Fang H, Zhou J, Wang Z, Qiu Z, Sun Y, Lin Y, Chen K, Zhou X, Pan M. Hybrid method integrating machine learning and particle swarm optimization for smart chemical process operations. *Frontiers of Chemical Science and Engineering*, 2022, 16(2): 274–287
  18. Ma Y, Gao Z, Shi P, Chen M, Wu S, Yang C, Wang J, Cheng J, Gong J. Machine learning-based solubility prediction and methodology evaluation of active pharmaceutical ingredients in industrial crystallization. *Frontiers of Chemical Science and Engineering*, 2022, 16(4): 523–535
  19. McGreavy C, Lu M, Wang X Z, Kam E K T. Characterisation of the behaviour and product distribution in fluid catalytic cracking using neural networks. *Chemical Engineering Science*, 1994, 49(24): 4717–4727
  20. Ochoa-Estopier L M, Jobson M, Smith R. Operational optimization of crude oil distillation systems using artificial neural networks. *Computers & Chemical Engineering*, 2013, 59: 178–185
  21. Yang F, Dai C, Tang J, Xuan J, Cao J. A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance. *Chemical Engineering Research & Design*, 2020, 155: 202–210
  22. Song W, Mahalec V, Long J, Yang M, Qian F. Modeling the hydrocracking process with deep neural networks. *Industrial & Engineering Chemistry Research*, 2020, 59(7): 3077–3090
  23. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324
  24. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*. Lille, France: JMLR, 2015
  25. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90
  26. Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: IEEE, 2015
  27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE, 2016
  28. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao R X. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 2019, 115: 213–237
  29. Serin G, Sener B, Ozbayoglu A M, Unver H O. Review of tool condition monitoring in machining and opportunities for deep learning. *International Journal of Advanced Manufacturing Technology*, 2020, 109(3): 953–974
  30. Souza R M, Nascimento E G, Miranda U A, Silva W J, Lepikson H A. Deep learning for diagnosis and classification of faults in industrial rotating machinery. *Computers & Industrial Engineering*, 2021, 153: 107060
  31. Yuan J, Tian Y. A multiscale feature learning scheme based on deep learning for industrial process monitoring and fault diagnosis. *IEEE Access: Practical Innovations, Open Solutions*, 2019, 7: 151189–151202
  32. Elhefnawy M, Ragab A, Ouali M S. Fault classification in the process industry using polygon generation and deep learning. *Journal of Intelligent Manufacturing*, 2022, 33(5): 1531–1544
  33. Glaeser A, Selvaraj V, Lee S, Hwang Y, Lee K, Lee N, Lee S, Min S. Applications of deep learning for fault detection in industrial cold forging. *International Journal of Production Research*, 2021, 59(16): 4826–4835
  34. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations*. San Diego, CA: OpenReview.net, 2015
  35. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, 2017
  36. Zagoruyko S, Komodakis N. Wide residual networks. In: *Proceedings of the British Machine Vision Conference (BMVC)*. York, UK: BMVA, 2016