Research Article

Analytical Redundancy of Variable Cycle Engine Based on Proper Net considering Multiple Input Variables and the Whole Engine's Degradation

Zihao Zhang , Xianghua Huang , and Tianhong Zhang

Nanjing University of Aeronautics and Astronautics, Jiangsu Province Key Laboratory of Aerospace Power System, Nanjing 210016, China

Correspondence should be addressed to Xianghua Huang; xhhuang@nuaa.edu.cn

Received 12 March 2021; Revised 15 April 2021; Accepted 26 April 2021; Published 27 May 2021

Academic Editor: Jiaqiang E

Copyright © 2021 Zihao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, Proper net is proposed to construct variable cycle engine's analytical redundancy, when all control variables and environmental variables change simultaneously, also accompanied with the whole engine's degradation. In another word, Proper net is proposed to solve a multivariable, strongly nonlinear, dynamic, and time-varying problem. In order to make the topological structure of Proper net physically explainable, Proper net's topological structure is designed according to physical relationship between variables, by which means analytical redundancy based on Proper net achieves higher accuracy with less calculation time. Experiments were compared with performance of analytical redundancy based on Proper net, seven convolutional neural network topological structures, and five shallow learning methods. Results demonstrate that under condition of average relative error less than 1.5%, Proper net is the most accurate and the least time-consuming one, which proves not only the effectiveness of Proper net but also the feasibility of topological structures' design method based on physical relationship.

1. Introduction

To guarantee the safety of aeroengines, diagnostics and fault-tolerance control have been developed [1]. For example, on premise of having installed mass flow sensors and pressure sensors at the entrance and exit of compressors, when aero-engine's rotation speed sensor fails, by interpolating compressor's characteristics picture using mass flow and pressure ratio, rotation speed could be uniquely identified. Thus, analytical redundancy is usually defined as signals calculated by algorithms or methods rather than signals measured by physical sensors. For diagnostics [2], analytical redundancy could be used as redundant signals for voting whether faults occur. In fault tolerance control [3], analytical redundancy could be used as alternative feedback signals of fault sensors so that the close loop of control system can keep intact and finish control task.

Methods of constructing analytical redundancy can be divided into model-based methods and data-driven algo-

rithms. Model-based analytical redundancy has been proposed for a long time and theoretically mature, whose representatives are Kalman filter and other improved Kalman filters. However, in practice, strong nonlinearity still makes Newton Raphson iteration or Euler iteration diverge as signals change rapidly or working points are away from the common operating line, which brings unsafe factors to model-based analytical redundancy [4]. This problem could be solved by constraining changing rate of environmental variables and control variables, but it is paradoxical to the emerging requirement—high mobility of military planes. Besides, in order to make model-based analytical redundancy more accurate, when aeroengines operate within a wide range of flight envelope, numerous sets of modification coefficients should be used to modify the onboard math model of aeroengines [5].

To avoid these model-related problems, scientists break a new path—data-driven analytical redundancy, by which means analytical redundancy could be built from historical

data instead of priori knowledge. Zhao and Sun proposed a Support Vector Machine (SVM) to construct analytical redundancy [6]. By adopting greedy stagewise and iterative strategies, the SVM is capable of online estimating parameters of complicated systems. Zhou et al. proposed an ELM to construct analytical redundancy for sensor fault diagnostics [1]. By selectively updating output weights of neural networks according to prediction accuracy and norms of output weight vectors, the prediction capability of the ELM is enhanced.

However, it can be seen that whether in Zhao's experiments or Zhou's simulations, the harshest situation is the dynamic process involving only three control variables, two components' degradation, and always at 0 height, 0 Mach number. A question is naturally proposed—why do not these methods consider more control variables, more points in flight envelope, and the whole engine's degradation? It is because SVM and ELM belong to shallow learning methods, a category of methods that do not have enough nonlinear expressive capability, which also explains why most shallow learning methods have to be online or onboard, because their parameters have to be updated to adapt to aeroengine's nonlinear characteristics at different working points.

Long-Short Time Memory neural network (LSTM) and convolutional neural network (CNN) are the most commonly used deep learning algorithms to solve strongly nonlinear problems. On one side, forget-and-memory gate helps increase nonlinear expressive capability of LSTM [7]. On the other side, forget-and-memory gate narrows the application fields of LSTM to sequence problems, like natural language processing where the input is a row vector. CNN originates from image processing problems, but widely used in language processing, classification, and regression problems. Babu et al. used CNN to estimate remaining useful life of aeroengines [8]. Through deep architecture, learned features are higher-level abstract representation of low-level raw sensor signals. Furthermore, feature learning and RUL estimation are mutually enhanced by supervised feedback. In order to accurately calculate fuel savings after aeroengine washing, convolutional neural network is used in Cui et al.'s research [9]. The results demonstrate that prediction accuracy gets improved by replacing integral operation with convolution operation. In Gou et al.'s research, a convolutional neural network (CNN) model trained with preprocessed and labeled data sets is used to extract features of a time-frequency graph, based on which faults can be identified and isolated [10].

However, whether in Babu et al.'s research, Cui et al.'s research, or Gou et al.'s research, design of CNN's topological structures is physically unexplainable, which makes CNN a black-box to practical physical problems [11]. On one side, being physically unexplainable may cause potential safe hazard to safety-critical objects, like aeroengines and cars. For example, if mapping relationship from input to analytical redundancy has an undetectable peak due to black-box property of CNN, and the peak analytical redundancy signal is used as reconstruction signal in aeroengines' fault-tolerance control, then the control loop could be unstable or collapse, which may lead to the whole engine's breakdown. Besides,

without a solid and explainable theory to support the design of topological structures applied in actual physical problems, many useless topological structures and huge amounts of redundant parameters would be kept, which means unnecessary computing overhead or lower accuracy. To be noted, during the process of applying all kinds of classical topological structures to analytical redundancy, such as Mobile net and Dense net, low accuracy and slow calculation speed are exactly the biggest obstacles that authors met.

Hence, on the basis of considering physical relationship between variables, an explainable topological structure named Proper net is designed to construct variable cycle engine's (VCE) analytical redundancy when all control variables and environmental variables change simultaneously, also accompanied with the whole engine's degradation. As for why VCE is chosen as research object and why the whole engine's degradation, control variables, and environmental variables are considered, it is only to make the problem more challenging, specifically making the problem in this paper a multivariable, strongly nonlinear, dynamic, and time-varying problem. Section 2 introduces the multivariable dynamic degradation data set and Proper net's structure. Following this, three experiments are conducted to demonstrate superiority of analytical redundancy based on Proper net in Section 3. Results are also discussed in this section. Section 4 concludes the paper tersely.

2. Method

2.1. VCE's Multivariable Dynamic Degradation Data Set. Figure 1 is VCE's structure diagram used in this paper. Authors refer to the two-bypass VCE's structure and the math model proposed in Aygun and Turan's paper [12], which has been validated and referred to by many other researchers. Due to the focus of this paper is not modeling of VCE, specific mathematic formulas have not been repeated in this paper. Main components of VCE include inlet (Inl), fan (fan), core driven fan (cdf), high-pressure compressor (hpc), combustion (cbt), high-pressure turbine (hpt), low-pressure turbine (lpt), mixer (Mix), nozzle (Noz), and bypass (Bps). Besides, mode switch valve (Msv), forward variable bypass ejector (Fvbe), and back variable bypass ejector (Bvbe) are used to change flow area. If the opening of Msv and Fvbe turns to 0, then VCE's operating mode will switch from turbofan into turbojet.

Degradation coefficients are defined as follows:

$$\begin{aligned} D_{\text{com},m} &= \frac{m_{\text{com,degradation}}}{m_{\text{com,nominal}}}, \\ D_{\text{com},e} &= \frac{e_{\text{com,degradation}}}{e_{\text{com,nominal}}}, \end{aligned} \quad (1)$$

where $D_{\text{com},m}$ and $D_{\text{com},e}$ represent components' degradation coefficients for mass flow and efficiency, respectively. $m_{\text{com,degradation}}$ is a component's mass flow after degradation, and $m_{\text{com,nominal}}$ is the nominal mass flow before degradation. $e_{\text{com,degradation}}$ is a component's adiabatic efficiency after

FIGURE 1: Component-level structure of VCE.

degradation, and $e_{com,nominal}$ is nominal adiabatic efficiency before degradation.

In this paper, ten degradation coefficients are defined and act on VCE simultaneously— $D_{fan,m}$, $D_{fan,e}$, $D_{cdf,m}$, $D_{cdf,e}$, $D_{hpc,m}$, $D_{hpc,e}$, $D_{hpt,m}$, $D_{hpt,e}$, $D_{lpt,m}$, and $D_{lpt,e}$. Their ranges are all from 0.96 to 1.

The original data set sizing 27×100000 includes multi-variable dynamic degradation simulation data of VCE. There is one row of time series with interval of 0.02 s, 2 rows of environmental variables (H, M_a), 5 rows of control variables ($W_f, A_8, M_{sv,ope}, F_{vbe,ope}, B_{vbe,ope}$), and 19 rows of state variables ($T_1, P_1, T_2, P_2, T_{21}, P_{21}, T_{22}, P_{22}, T_7, P_7, T_8, P_8, P_5, T_6, N_l, N_h, P_3, T_5, P_6$). All variables are listed in Table 1. Rows from 2 to 22 are used as input data, and N_l, N_h, P_3, T_5, P_6 are variables to be estimated. In this paper, we only take N_l estimation as example.

As shown in Figure 2, a data map sizing 21×42 from 2nd row (t_{m1}) to 22nd row (t_{m42}) is firstly segmented from original data set, and then, the data map is duplicated and reassembled into an input data map sizing 42×42 . Corresponding output of the input data map is N_l at t_{m42} moment. Other input data maps are also made in the same way. After segmentation and reassembly, 99959 input data map sizing 42×42 are generated from the original data set. Corresponding output of each input data map is N_l from 42rd column to 100000th column, amounting to 99959 columns. To be clearer, 99959 data maps are segmented from 1-42rd column, 2-43th column, 3-44th column,, 99959-100000th column. Among all data maps, 79959 data maps are used for training, 10000 data maps for validation, and 10000 data maps for test.

Traditional ELM or SVM only uses 2 or 3 control period's information before the estimated moment, such as Zhao and Sun's research [6] and Zhou et al.'s research [1]. However, delay caused by burning, moment of inertia or large-volume mixer is usually longer than two or three control periods, which implies massive historical information has not been efficiently utilized.

The method proposed in this paper needs 21 input variables, which is obviously different from traditional methods. In many researches, only 5 or 6 sensors are used as inputs, but this is exactly why authors argue traditional methods cannot solve such a complicated problem proposed in this paper. No matter how excellent the traditional methods are, the information that can be utilized is far too less that it cannot help the judgement on the whole engine's degradation,

not to mention constructing analytical redundancy under condition of degradation and multiple variables. What needs to be noted is that number of sensors used in Zedda and Singh's research has arrived at 16 in 2002, which means number of sensors deployed in advanced aeroengines at present or in the future has reached up to or even exceeds 21 [13].

Reasons why data maps should be segmented and reassembled like Figure 2 are listed as follows:

- (1) Data maps are made into quadrate for compatibility. In classical CNNs, input maps are usually quadrate, but with the development of CNN, rectangle inputs and kernels are introduced, which means that rectangle inputs are not compatible with some classical CNNs, like Alex net used in this paper
- (2) Duplication of original data helps improve accuracy and calculation speed. From the perspective of connectionism, connection between rows may form key features to increase accuracy, but it is not known in advance which feature is more important. The general approach is directly connecting as many rows as possible. Input rows of original data are from 2 to 22, and it is tricky to build connection between row 2 and row 22 unless using a kernel sizing 21×21 . If so, the number of weights and biases will increase steeply, also accompanied slower calculation speed. Thus, a relatively less time-consuming duplication is performed. As a result, original data set sizing 21×100000 is transformed into 42×100000 , and connection between any two rows can be made with kernel only sizing 11×11
- (3) Although the number of columns used in data maps is not analytic, it has to keep information sufficient. To a great extent, useless information will be automatically filtered by adjusting weights and biases during process of backpropagation. Besides, different from simple application situations, multivariable dynamic degradation analytical redundancy needs more significant information hidden in data to form deep features

2.2. Net Structure and Training Options. Feature extraction capability is directly related with macroscopic topological structures. Compared to Alex net [14] with only one path, Dense net's macroscopic topological structure features two

TABLE 1: Explanation of variables in original data set.

Row	Symbol	Explanation	Unit	Range
1	t	Time	s	0~2000
2	H	Flight height	km	0~5
3	M_a	Flight Mach number		0~1
4	W_f	Fuel consumption	kg/s	2~2.5
5	A_8	Nozzle area	m ²	0.2~0.25
6	$M_{sv,ope}$	The opening of Msv		75~100
7	$F_{vbe,ope}$	The opening of forward variable bypass ejector		75~100
8	$B_{vbe,ope}$	The opening of backward variable bypass ejector		75~100
9	T_1	Fan inlet total temperature	K	
10	P_1	Fan inlet total pressure	Pa	
11	T_2	Core driven fan inlet total temperature	K	
12	P_2	Core driven fan inlet total pressure	Pa	
13	T_{21}	High-pressure compressor inlet total temperature	K	
14	P_{21}	High-pressure compressor inlet total pressure	Pa	
15	T_{22}	Bypass outlet total temperature	K	
16	P_{22}	Bypass outlet total pressure	Pa	
17	T_7	Mixer outlet total temperature	K	
18	P_7	Mixer outlet total pressure	Pa	
19	T_8	Nozzle outlet total temperature	K	
20	P_8	Nozzle outlet total pressure	Pa	
21	P_5	High-pressure turbine outlet total pressure	Pa	
22	T_6	Low-pressure turbine outlet total temperature	K	
23	N_l	Low-pressure rotor speed	rpm	
24	N_h	High-pressure rotor speed	rpm	
25	P_3	High-pressure compressor outlet total pressure	Pa	
26	T_5	High-pressure turbine outlet total temperature	K	
27	P_6	Low-pressure turbine outlet total pressure	Pa

FIGURE 2: Segmentation and reassembly of the original data set.

paths [15], which greatly improves Dense net's accuracy. In addition, on the basis of Dense net's macroscopic topological structure, Mobile net's macroscopic topological structure only repeats convolution layer and batch normalization layer one more time [16]. As a result, the number of layers used in Mobile net reduces to around one-fourth as many as that in Dense net, and Mobile net's calculation time also declines dramatically. These examples are just parts of the function of CNN's topological structures, but it can be seen CNN's topological structure has a great impact on its performance.

The topological structure of Proper net is demonstrated in Figure 3. The whole structure can be divided into three levels, and the first level is Proper inner-level data fusion (PINDF). The second level is referenced to the classical inter-level data fusion structure—inception net [17]. The third level is an ordinary cascade structure consisting of convolutional layers, activation layers and normalization layers. At last, the whole net ends with a succession of two fully connected layers, and mean square error (MSE) is used as loss function.

What needs to be specified is that most hyperparameters are empirical rather than analytic. Hyperparameters refer to those parameters that are unable to be automatically modified, such as topological structure, choice of activation layer, the number of kernels, and neural nodes. In this paper, the number of kernels, kernel stride, the method of normalization, batch size, activation layer, and neural nodes of fully connected layer are set to 32, 1 * 1, batch normalization, 128, LeakyRelu, 100.

2.2.1. Convolution Layer [18]. The input of convolution layer is convolved with several convolution kernels, and after that, the result will be added with a bias. The output data map is calculated as follows:

$$x_j^l = \sum_i x_i^{l-1} * k_{ij}^l + s_j^l, \quad (2)$$

where $*$ denotes a convolution operator, and x_j^{l-1} and x_j^l are input and output of convolution layers, standing for the i_{th} channel of $l-1_{th}$ layer and the j_{th} channel of l_{th} layer separately. k_{ij}^l means kernels of l_{th} layer, related with x_j^{l-1} and x_j^l , and s_j^l represents the bias of j_{th} channel of l_{th} layer.

2.2.2. Batch Normalization Layer [19]. Batch normalization layer can prevent gradient explosion or disappearance, improve robustness of a model, and keep activation function away from its saturated region.

$$\mu_{i,j}^{l-1} = \frac{1}{b} \sum_{t=1}^b x_{i,j}^{l-1,t},$$

$$\sigma_{i,j}^{l-1} = \sqrt{\frac{1}{b} \sum_{t=1}^b \left(x_{i,j}^{l-1,t} - \mu_{i,j}^{l-1} \right)^2 + \epsilon_{i,j}^{l-1}},$$

$$\hat{x}_{i,j}^l = \frac{x_{i,j}^{l-1} - \mu_{i,j}^{l-1}}{\sigma_{i,j}^{l-1}},$$

$$x_{i,j}^l = \gamma_{i,j}^l \cdot \hat{x}_{i,j}^l + \beta_{i,j}^l, \quad (3)$$

where b is the batch size, $\mu_{i,j}^{l-1}$ and $\sigma_{i,j}^{l-1}$ are mean value and corrected standard deviation of $x_{i,j}^{l-1,t}$, which represents j_{th} value of i_{th} channel in $l-1_{th}$ layer of t_{th} data map within a batch. $\epsilon_{i,j}^{l-1}$ is the correction factor defined as 0.0001 for steady training. $\gamma_{i,j}^l$ and $\beta_{i,j}^l$ are regulatory factors to be learned.

2.2.3. Activation Layer. This paper uses LeakyRelu as activation layer [20], whose formula is given below.

$$x_{i,j}^l = \begin{cases} x_{i,j}^{l-1} & \text{if } x_{i,j}^{l-1} > 0, \\ 0.002 \cdot x_{i,j}^{l-1} & \text{if } x_{i,j}^{l-1} < 0, \end{cases} \quad (4)$$

where $x_{i,j}^{l-1}$ and $x_{i,j}^l$ are input and output element of the activation layer, standing for j_{th} element in i_{th} channel of $l-1_{th}$ layer and the corresponding element of l_{th} layer separately.

2.2.4. Addition Layer. Addition layer adds inputs by element.

$$x_{i,j}^l = x_{i,j}^{l-1} + x_{i,j}^{l-2} + \dots + x_{i,j}^{l-m} \quad m > 1, \quad (5)$$

where m is the number of inputs.

2.2.5. Depth Concatenation Layer. Depth concatenation layer connects all inputs along channels.

$$x^l = x^{l-1} \oplus x^{l-2} \oplus \dots \oplus x^{l-m} \quad m \geq 1, \quad (6)$$

where $\oplus$ is the operator connecting inputs along channels. For example, after $\oplus$ operator, two inputs with 20 channels and 30 channels separately will be combined into an output with 50 channels.

2.2.6. Fully Connected Layer. For fully connected layer, all nodes are connected with each other.

$$x_j^l = \sum_i x_i^{l-1} \cdot w_{ij}^l + s_j^l, \quad (7)$$

where x_i^{l-1} and x_j^l are i_{th} node of $l-1_{th}$ layer and the j_{th} node of l_{th} layer, respectively. w_{ij}^l means weights of l_{th} layer, related with x_i^{l-1} and x_j^l . s_j^l represents the bias of j_{th} node of l_{th} layer.

2.2.7. Regression Layer. Regression layer defines the loss function. MSE is used in this paper.

$$L = \frac{1}{m} \sum_{i=1}^m (y_i - y_{i,a})^2, \quad (8)$$

where L is loss. y_i and $y_{i,a}$ are estimation output value and actual output value. m is the number of outputs.

FIGURE 3: Proper net structure.

FIGURE 4: Comparison between Proper inner-level data fusion structure (PINDF) and over inner-level data fusion structure (OINDF).

TABLE 2: Training options.

Optimizer	Learning strategy	Total epochs	Initial learning rate	Software environment
Adam	Piecewise	10	0.0001	Matlab 2019a
Batch size	LearnRateDropFactor	LearnRateDropPeriod	Shuffle	GPU
128	0.1	2	Every epoch	RTX 2080Ti 11G

Topological structure of CNN often includes hundreds of layers, so it is inconvenient to analyze relationship between the whole net and CNN's performance. Usually, one or two basic macroscopic topological structures will be abstracted from the whole net, which means a novel CNN structure is always accompanied with a new macroscopic structure. For Proper net, although its second and third level are referenced to inception net and other nets, its first level—Proper inner-level data fusion structure—is totally original and

the whole topological structure is unique. In Figure 4, as the macroscopic structure and highlight of Proper net, PINDF proposed in this paper is demonstrated and compared with macroscopic structure of Google net named OINDF [21].

As shown in Figure 4, PINDF takes advantage of six kinds of kernel sizes to extract features in different scales. In many image-processing problems, requirement for real-time calculation is not as strict as that in aeroengines. So usually, two or

TABLE 3: Performances of different convolutional neural networks.

Net	Average relative error/%	Layers	Training time/min (10 epochs)	Calculation time/s (1000 points)	Macroscopic structure
Proper net	0.84 _{1st}	49 _{1st}	25 _{1st}	0.3645 _{1st}
Dense net	1.17	707	535	1.4221
Mobile net	1.25	153	73	0.3716
Res101	1.33	143	155	0.9655
Alex net	12.71	24	6	0.1284
Vgg19 net	7.43	40	26	0.5133

TABLE 3: Continued.

Net	Average relative error/%	Layers	Training time/min (10 epochs)	Calculation time/s (1000 points)	Macroscopic structure
Google net	2.05	143	40	0.2317
Squeeze net	3.85	68	17	0.1192

Yellow rectangle: convolution layer; red rectangle: activation layer; green rectangle: pooling layer; violet rectangle: activation layer; white rectangle: fusion layer.

Proper Inner-Level Data Fusion (PINDF)

Proper Inner-Level Data Fusion (PINDF)

TABLE 4: Comparison between over inner-level data fusion structure and interlevel data fusion structure.

Net	Average relative error/%	Layers	Training time/min (10 epochs)	Calculation time/s (1000 points)	Macroscopic structure
Proper net	0.84	49	25	0.3645	
Google-alter net	1.74	31	12	0.2117	
Mobile-alter net	1.22	63	52	0.3694

FIGURE 5: Continued.

FIGURE 5: N_I estimation in dynamic process of all eight convolutional neural networks.

TABLE 5: The list of experiment conditions.

(a)

	H/km	M_a	$W_f/(\text{kg} \times \text{s})$	A_g/m^2	$M_{\text{sv,ope}}$	$F_{\text{vbe,ope}}$	$B_{\text{vbe,ope}}$	Status
Point 1 (P1)	0	0	2	0.2	75	75	75	Steady
Point 2 (P2)	5	1	2.5	0.25	100	100	100	Steady
Small range 1 (SR1)	0-0.1	0-0.1	2-2.1	0.2-0.21	75-80	75-80	75-80	Steady
Small range 2 (SR2)	2.5-2.6	0.5-0.6	2.25-2.35	0.225-0.235	87.5-92.5	87.5-92.5	87.5-92.5	Dynamic
Wide range 1 (WR1)	0-5	0-1	2-2.5	0.2-0.25	75-100	75-100	75-100	Steady
Wide range 2 (WR2)	0-5	0-1	2-2.5	0.2-0.25	75-100	75-100	75-100	Dynamic

(b)

	$D_{\text{fan},m}$	$D_{\text{fan},m}$	$D_{\text{fan},m}$	$D_{\text{fan},m}$	$D_{\text{fan},m}$
Point 1 (P1)	0.96	0.96	0.96	0.96	0.96
Point 2 (P2)	1	1	1	1	1
Small range 1 (SR1)	0.96-0.98	0.96-0.98	0.96-0.98	0.96-0.98	0.96-0.98
Small range 2 (SR2)	0.98-1	0.98-1	0.98-1	0.98-1	0.98-1
Wide range 1 (WR1)	0.96-1	0.96-1	0.96-1	0.96-1	0.96-1
Wide range 2 (WR2)	0.96-1	0.96-1	0.96-1	0.96-1	0.96-1
Point 1 (P1)	0.96	0.96	0.96	0.96	0.96
Point 2 (P2)	1	1	1	1	1
Small range 1 (SR1)	0.96-0.98	0.96-0.98	0.96-0.98	0.96-0.98	0.96-0.98
Small range 2 (SR2)	0.98-1	0.98-1	0.98-1	0.98-1	0.98-1
Wide range 1 (WR1)	0.96-1	0.96-1	0.96-1	0.96-1	0.96-1
Wide range 2 (WR2)	0.96-1	0.96-1	0.96-1	0.96-1	0.96-1

three small sizes of kernels are used to extract microscopic features, and then, through hundreds or even thousands of layers, these features are recombined to classify different objects. However, for aeroengines, microscopic and macroscopic features should be extracted directly and concisely to decrease the number of layers for saving calculation time.

So, six kinds of kernel sizes are used in PINDF instead of three kernel sized like that in OINDF.

Considering physical relationship between variables, the lower bound and upper bound of these kernels have been decided in Section 2.1, which are 1×1 and 11×11 , respectively. Of course, kernel sizing from 2×2 to 12×12 can also

TABLE 6: The ARE of six methods.

%	P1	P2	SR1	SR2	WR1	WR2
Proper net	0.0001 _{eq}	0.001 _{eq}	0.43 _{1st}	0.58 _{1st}	0.52 _{1st}	0.84 _{1st}
SVM	0.0001	0.001	0.68	1.21	13.34	15.37
DT	0.0001	0.0001	2.09	3.82	18.64 _{6th}	27.98 _{6th}
KNN	0.0001	0.0001	2.66 _{6th}	4.94 _{6th}	14.96	24.11
FC	0.0001	0.0001	1.23	1.56	9.74	11.56
LSTM	0.0001	0.0001	0.67 _{2nd}	0.81 _{2nd}	6.98 _{2nd}	8.33 _{2nd}

TABLE 7: N_f ARE of different sizes of data maps.

%	Proper net	Dense net	Mobile net	Res101 net
21 * 21	1.21	1.44	1.68	1.84
42 * 42	0.84	1.17	1.25	1.33
63 * 63	0.84	1.16	1.26	1.33

achieve similar effects, but the price is more calculation time and storage space.

Another feature of PINDF is that six paths are combined by pair instead of getting combined together. On one side, if all paths are combined together, different scales of features will be mixed together. During the process of back propagation, it would become harder to adjust kernels' weights to differentiate importance of different scales of features. On the other side, it can be seen that kernels are combined equidistantly with 6, for example, $1 * 1 - 7 * 7, 3 * 3 - 9 * 9, 5 * 5 - 11 * 11$. From the perspective of physical meanings, the larger the kernel sizes are, the more the number of variables and the sampling moments are. Smaller kernels perform reversely. When one or two variables change, while the other variables keep static, changing magnitude of features formed by small kernels is greater, which means output is more sensitive to smaller kernels. However, if the whole engine degrades where at least ten variables are changing together, larger kernels are dominant in output. In order to make full use of advantages of smaller and larger kernels and save calculation time, paths are combined equidistantly by pair.

3. Experiments, Results, and Discussion

3.1. Proper Net's Top Accuracy and Fastest Calculation Speed. Seven classical topological structures that are used in image processing are modified appropriately to adapt to analytical redundancy, like Res101 net [22], Squeeze net [23], and Vgg19 net [24]. This paper changes classical nets' input size, number of output nodes, and output layer into $42 * 42$, 1 and regression layer, but keeps their macroscopic topological structures that are critical to deep feature extraction unchanged.

All nets are trained with options listed in Table 2. The learning strategy is piecewise, in our case, which means the learning rate will be multiplied by 0.1 (LearningRateDropFactor) every 2 epochs (LearningRateDropPeriod). To increase fitting accuracy and generation ability, all data maps in training will be shuffled every epoch. Besides, considering changing learning rate, we decide to use Adaptive moment

estimation (Adam) to update weights and biases [25]. The software and hardware environment of training are Matlab 2019a and RTX2080Ti 11G separately.

Table 3 demonstrates the performance of multivariable dynamic degradation analytical redundancy based on eight kinds of macroscopic topological structures. Simulation results are based on test data set that is made in Section 2.1, and N_f is picked as analytical redundancy. For aeroengines, tolerant ARE is around 1.5%, which means those nets with ARE higher than 1.5% cannot be used to aeroengines' analytical redundancy. Among all nets whose ARE lower than 1.5%, Proper net has the lowest ARE (0.81%), minimum number of layers (49), the least training time (25 minutes every 10 epochs), and fastest calculation speed (0.3645 second every 1000 points). All of the four indicators prove that the performance of analytical redundancy based on Proper net is better than that based on other nets.

In Table 3, it can be seen that nets whose ARE is lower than 1.5% have structures similar to Proper interlevel data fusion (PITDF), while nets whose ARE is higher than 1.5% do not have. Thus, it can be concluded PINDF helps improve the accuracy of aeroengine's analytical redundancy. Besides, among all nets whose ARE is lower than 1.5%, Proper net is the only net that has both PINDF and PITDF, while Mobile net and Dense net only have PITDF, which is why Proper net's performance is better than Mobile net and Dense net. Results demonstrate that PINDF could extract deep features of aeroengines and further improve performance of analytical redundancy.

In Table 4, PINDF and PITDF of Proper net are altered with OINDF used in Google net and over interlevel data fusion (OITDF) used in Mobile net. Nets after alteration are named as Google-alter net and Mobile-alter net, respectively. Simulation is based on the multivariable dynamic degradation test data set made in Section 2.1, and analytical redundancy is N_f . After alteration, Proper net's accuracy deteriorates dramatically. ARE of Google-alter net is approximately two times as much as that of Proper net. Although Mobile-alter net's accuracy is still lower than 1.5%, its training time is 27 minutes more than Proper net's training time. Therefore, it can be ascertained that OINDF and OITDF would reduce the accuracy of analytical redundancy and increase training time.

3.2. Dynamic Performance of Multivariable Dynamic Degradation Analytical Redundancy. To validate the effectiveness of Proper net, experiments under conditions of more than 100 stable points and 70 kinds of dynamic process have been conducted, which include small-step responses, large-step responses, slope responses, sigmoid responses, single variable, two or three variables, one or two components' degradation, and the whole engines' degradation. All results of these experiments can meet the requirement of accuracy and calculation speed. Due to the space limitation of the paper, the most representative and the harshest situation is picked up to display performance of Proper net. In this situation, seven step signals— $H(0 \sim 5 \text{ km})$, $M_a(0 \sim 1)$, $W_f(2 \sim 2.5 \text{ kg/s})$, $A_g(0.2 \sim 0.25 \text{ m}^2)$, $M_{sv,ope}(100 \sim 75)$, $F_{vbe,ope}(100 \sim 75)$, and $B_{vbe,ope}(100 \sim 75)$ —act on the VCE

simultaneously at 5 s while all ten degradation coefficients change from 1 to 0.96. Although, in practice, it is not allowed to make all the seventeen variables change together by such big steps, and according to experience, if an algorithm can overcome the harshest situation, performance of analytical redundancy based on Proper net should be better in some simple situations, like slope signals, small-magnitude steps, or just two or three signals changing at the same time.

In Figure 5, red lines are actual N_I , and green lines are estimated value of N_I .

As shown in Figure 5, when step signals act on VCE at 5 s, Proper net responses immediately, but for other nets, there is an obvious delay between actual N_I and estimation N_I . Besides, it can be seen that both of dynamic performance and steady performance of analytical redundancy based on Proper net are the best. This is because Proper net is designed by considering physical relationship between aeroengine's variables while other nets are designed for image processing problems.

3.3. The Superiority of Proper Net to Other Data-Based Methods. In order to demonstrate the superiority of Proper net, other five kinds of data-based methods are compared with Proper net under six conditions. Other data-driven methods used to estimate N_I include SVM, Decision Tree (DT) [26], K Nearest Neighbor (KNN) [27], Fully Connected neural network (FC) [28], and Long-Short Memory neural network (LSTM) [29]. Six conditions are listed in Table 5. For point 1 (P1) and point 2 (P2), all environment variables and control variables are set to constants. For P1, its training data and test data are the same point, and P2 is also trained and tested in the same way. Dynamic status means that data are sampled during dynamic process in the way described in Section 2.1. Ranges of wide range 1 (WR1) and wide range 2 (WR2) are wider than those of SR1 and SR2. The experiment conditions are listed in Table 5.

The AREs of all six methods under six conditions are demonstrated in Table 6. AREs of analytical redundancy based on Proper net are less than any other methods under any conditions. The results also show that as far as accuracy is concerned, analytical redundancy based on Proper net outperforms analytical redundancy based on other methods.

In Table 6, as the conditions change from P1 to WR2 in the order from left to right, the difficulty of estimation ascends gradually. For P1 and P2, AREs of analytical redundancy based on Proper net are similar to those based on other methods, but with the condition gets harsher and harsher, the gap of AREs between Proper net and other methods dramatically increases. For SR1, the gap of AREs between analytical redundancy based on the lowest-ARE method and the highest-ARE method is 2.23%, while when it comes to WR2, the gap turns out to be 27.14%, which increases by more than ten times. It can be concluded that as a deep learning method, Proper net proposed in this paper performs better than other shallow learning methods.

3.4. Comparison between Different Sizes of Input Data Maps. To quantitatively validated the effectiveness of input data map's segmentation and reorganization, data maps sizing

21 * 21, 42 * 42, 63 * 63 are tested by Proper net, Mobile net, Dense net, and Res101 net with N_I test data sets. 21 * 21 data maps mean 21 features along with 21 sample periods. 42 * 42 data maps that include 42 sample periods are made with a duplication alongside row direction. Data map sizing 63 * 63 is made from original data maps sizing 21 * 63 with two times of duplication by row direction. All sizes of kernels are in consistency with those used in Section 3.1.

Table 7 demonstrates N_I AREs of three sizes of data maps tested with four nets. It can be seen that AREs of 42 * 42 data maps tested with four nets are all lower than that of data map sizing 21 * 21. When the size of data maps expands from 42 * 42 to 63 * 63, the ARE basically remains static, only with a slight decrease by 0.01% for Dense net and a marginal rise by 0.01% for Mobile net.

In practical physical process, features in top rows and bottom rows of data maps function together to N_I , but as mentioned in Section 2.2, 21 * 21 data maps cannot build direct relationship between top rows and bottoms rows. The disconnection is why 21 * 21 data maps perform worse than 42 * 42 data maps. Due to data map sizing 63 * 63 will not build any new direct connection between variables, it is reasonable that the accuracy of 63 * 63 data maps basically is the same with data map sizing 42 * 42.

4. Conclusions

- (1) A novel convolutional neural network topological structure named Proper net is proposed to construct variable cycle engine's analytical redundancy when all control variables and environmental variables change simultaneously, also accompanied with the whole engine's degradation
- (2) To sufficiently utilize the underlying information of aeroengine's sensors, original sensor data is segmented and reassembled into data maps that contain more historical information
- (3) On the basis of considering physical relationship between aeroengine's variables, Proper inner-level data fusion structure is specially designed to improve the accuracy and calculation speed of aeroengine's analytical redundancy
- (4) Compared to shallow learning methods and other convolutional neural network structures used for image processing, Proper net owns highest accuracy and fastest calculation speed

Data Availability

The data set used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant numbers: 51576097 and 51976089) and Foundation Strengthening Project of the Military Science and Technology Commission (Grant number: 2017-JCJQ-ZD-047-21).

References

- [1] J. Zhou, Y. Liu, and T. Zhang, "Analytical redundancy design for aeroengine sensor fault diagnostics based on SROS-ELM," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8153282, 9 pages, 2016.
- [2] J. Lu, J. Huang, and F. Lu, "Sensor fault diagnosis for aero engine based on online sequential extreme learning machine with memory principle," *Energies*, vol. 10, no. 1, p. 39, 2017.
- [3] S. Anwar and L. Chen, "An analytical redundancy-based fault detection and isolation algorithm for a road-wheel control subsystem in a steer-by-wire system," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2859–2869, 2007.
- [4] J. Lu, Y. Guo, and S. Zhang, "Aeroengine on-board adaptive model based on improved hybrid Kalman filter," *Journal of Aerospace Power*, vol. 26, no. 11, pp. 2593–2600, 2011.
- [5] S. PANG, Q. Li, H. Feng, and H. Zhang, "Joint steady state and transient performance adaptation for aero engine mathematical model," *IEEE Access*, vol. 7, pp. 36772–36787, 2019.
- [6] Z. Yongping and S. Jianguo, "Fast online approximation for hard support vector regression and its application to analytical redundancy for aeroengines," *Chinese Journal of Aeronautics*, vol. 23, no. 2, pp. 145–152, 2010.
- [7] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.
- [8] G. S. Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *International Conference on Database Systems for Advanced Applications*, pp. 214–228, Springer, 2016.
- [9] Z. Cui, S. Zhong, and Z. Yan, "Fuel savings model after aero-engine washing based on convolutional neural network prediction," *Measurement*, vol. 151, 2020.
- [10] L. Gou, H. Li, H. Zheng, H. Li, and X. Pei, "Aeroengine control system sensor fault diagnosis based on CWT and CNN," *Mathematical Problems in Engineering*, vol. 2020, 12 pages, 2020.
- [11] N. Wang, M. Chen, and K. P. Subbalakshmi, "Explainable CNN-attention networks (C-attention network) for automated detection of Alzheimer's disease," 2020, <http://arxiv.org/abs/2006.14135>.
- [12] H. Aygun and O. Turan, "Exergetic sustainability off-design analysis of variable-cycle aero-engine in various bypass modes," *Energy*, vol. 195, 2020.
- [13] M. Zedda and R. Singh, "Gas turbine engine and sensor fault diagnosis using optimization techniques," *Journal of Propulsion and Power*, vol. 18, no. 5, pp. 1019–1025, 2002.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, Honolulu, HI, USA, 2017.
- [16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, Salt Lake City, UT, USA, 2018.
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Las Vegas, NV, USA, 2016.
- [18] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, <http://arxiv.org/abs/1404.2188>.
- [19] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, <http://arxiv.org/abs/1502.03167>.
- [20] X. Zhang, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, London, UK, 2017.
- [21] P. Ballester and R. Araujo, "On the performance of GoogLeNet and AlexNet applied to sketches," in *Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA, 2016.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [23] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, <http://arxiv.org/abs/1602.07360>.
- [24] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [25] D. P. Kingma and J. BA, "Adam: a method for stochastic optimization," 2014, <http://arxiv.org/abs/1412.6980>.
- [26] Y. Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.
- [27] W. Dong, C. Moses, and K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures," in *Proceedings of the 20th international conference on World wide web*, pp. 577–586, Hyderabad, India, 2011.
- [28] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4580–4584, South Brisbane, QLD, Australia, 2015.
- [29] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2017.