

Remaining Useful Life Estimation for Turbofan Engine with Transformer-based Deep Architecture

Qianxia Ma *

Department of Automation
Tsinghua University
Beijing, China
mqx15@mails.tsinghua.edu.cn

Ming Zhang *

College of Engineering and Physical Sciences
Aston University
Birmingham, UK
m.zhang21@aston.ac.uk

Yuchun Xu

College of Engineering and Physical Sciences
Aston University
Birmingham, UK
y.xu16@aston.ac.uk

Jingyan Song

Department of Automation
Tsinghua University
Beijing, China
jysong@tsinghua.edu.cn

Tao Zhang [†]

Department of Automation
Tsinghua University
Beijing, China
taozhang@tsinghua.edu.cn

Abstract—With the development of information technology and sensors, the large industrial system has become a data-rich environment, which leads to the rapid development and application of deep learning for the remaining useful life prediction, especially for the turbofan engine. Currently, the deep architecture of CNN, LSTM have been used to address the RUL estimation of a turbofan engine. However, they are mainly focused on simulation degradation data. The new realistic run-to-failure turbofan engine degradation dataset has been published in 2021, which presents a significant difference from the simulation one. The main challenge is that the flight duration of each cycle is different, which will result in the current deep method hardly used for predicting the RUL for the practical degradation data. To tackle this challenge, we propose a novel Transformer-based model using guiding features to deal with the unfixed-length data. Besides, our G-Transformer model makes use of multi-head attention to access the global features from various representation subspaces. We conduct experiments on turbofan engine degradation data with variable-length input under practical flight conditions. Empirical results and feature visualization via t-SNE indicate the effectiveness of the G-Transformer model for RUL estimation of turbofan engines.

Index Terms—Remaining useful life, Turbofan engine, Deep learning, Attention mechanism, Transformer.

I. INTRODUCTION

As the critical technique of prognostic and health management(PHM), remaining useful life(RUL) estimation plays an important role in the predictive maintenance of large equipment, which aims at estimating the performance of machinery across its lifetime period, and providing a suitable maintenance plan to avoid serious accidents occurring. With the development of the Internet of Things(IoT) and industrial digitalization, the current industrial systems have been gradually turned to become a data-rich environment. These changes have created an unprecedented opportunity to research and

develop the advanced RUL method and application by the powerful deep learning technique [1].

A turbofan engine is the most critical power unit for the aircraft, and it is obvious that the safety and stable working of them are very important. However, the turbofan engine generally works in an extremely complex and harsh environment, which makes the need for predicting the RUL lifetime of them extremely urgent. Recently, there are many studies of predicting the RUL lifetime of turbofan engine based on powerful deep learning [2]–[7]. However, they are focus on the degradation simulation data of turbofan engine carried out using C-MAPSS [8]. There are significant differences between the simulation degradation data and the real-world degradation data because the practical engine data collect from the complex working condition, e.g., the flight duration of each cycle task is different, which has been presented in the new turbofan engine degradation dataset [9]. The current RUL method based on deep learning cannot deal with the realistic dataset of run-to-failure trajectories for the aircraft engines under realistic flight conditions because the duration of each flight task is not equal. We try to tackle this challenge with inspiration from sequence modeling, and a novel remaining useful life estimation method based on the transformer-based deep architecture has been proposed for the new degradation dataset of a turbofan engine.

II. RELATED WORKS

RUL estimation for turbofan engines is a very hot topic because its research and application are of great significance and value. However, many researchers have studied the advanced method and mainly testified on the simulation dataset since almost all the practical data of the turbofan engine are very precious and even confidential. Heimes et al. [2] first apply RNN to the degradation simulation data of turbofan engine. Xia et al. [3] proposed the hybrid deep RNN for predicting the remaining lifetime. Meanwhile, Li et al. [4] and Li et al. [5] proposed to use the deep CNN to address this

* Q. Ma and M. Zhang contribute equally to this work.

[†] Tao Zhang is the corresponding author.

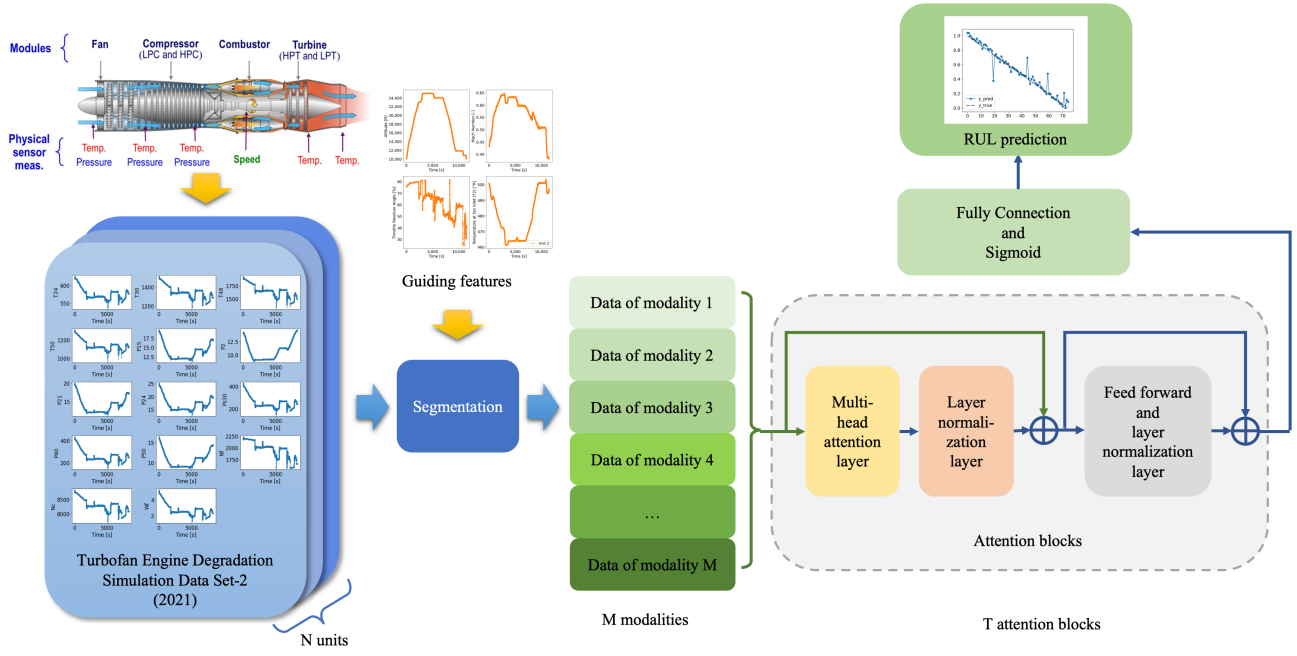


Fig. 1. The framework of G-Transformer model.

problem. Then, many studies combine the advantage of CNN and LSTM [10], [11], while others try to prediction the RUL from the sequence learning perspective [6] and attention-based deep learning approach [7]. These data-based methods have been performed very well on the simulation data scenarios.

Turboprop Engine Degradation Data Set-2(N-CMAPSS) [9] was published by the Prognostics CoE at NASA Ames in collaboration with ETH Zurich and PARC in 2021. Compared with the former Turboprop Engine Degradation Simulation Data Set(2008) [8], N-CMAPSS provides run-to-failure trajectories for a small fleet of aircraft engines under realistic flight conditions. Each unit consists of a various number of flight cycles of unfixed length.

To process variable-length input, models based on Transformer architecture are introduced in sequence modelling. The Transformer model is proposed by Google in 2017 [12]. Instead of commonly-used CNN and RNN, Transformer only makes use of multi-head attention to extract features. From then on, variants of the Transformer have gained substantial progress in Natural Language Processing(NLP) tasks and other domains. As one of the typical representatives, Bidirectional Encoder Representations from Transformers(BERT) is a pre-trained NLP model, which was also proposed by Google [13]. Benefit from two pretrained tasks, BERT surpasses state-of-the-art models on eleven natural language processing tasks. Models in the GPT [14] family are structured based on the encoder of the well-known Transformer model. In 2021, the third version of the Generative Pre-Training(GPT 3) model [15] has shown its incredible capability of handling numerous tasks within and outside the domain of NLP, such as question and

answering, calculation, article writing, and translation.

In this paper, we propose a novel remaining useful life estimation model for turboprop engines based on the transformer framework. Multihead-attention is adopted to extract features from variable-length input. Moreover, we enhance the model by introducing guiding features to segment modalities for better sampling.

III. METHODOLOGY

The G-Transformer model is mainly inspired by the encoder of the Transformer model in natural language processing. Figure 1 shows the whole framework of the G-transformer model. The model is composed of three parts, that are sampling, feature extraction, and prediction. Turboprop Engine Degradation Data Set-2 provides run-to-failure trajectories of different units. Scenario descriptors are used as the guiding features for modality segmentation. The embedding layer processes the high-dimensional input data and transmits the embedded data to the feature extraction blocks to acquire high-level semantic features. Data of different modalities are concatenated to form the input of feature extraction. The feature extraction blocks share similar architecture with the encoder of the Transformer model, which is mainly made up of multi-head attention layers, normalization layers, feed-forward layers, and residual connections. After repeating T attention blocks, the high-level features are extracted for further prediction. The remaining useful life prediction of a turboprop engine is regarded as a regression task. The features are sent to fully connected layers and Sigmoid function. As a result, the G-Transformer model

predicts a number between 0 and 1 to show the remaining useful life.

A. Sampling

Turbofan Engine Degradation Data Set-2 consists of data of several units. Compared with the former dataset in 2008, this new N-CMAPSS dataset is organized with units of cycles. More specifically, the dataset provides units. Within each unit, the flight data has flight cycles with various flight lengths. Table I shows three flight classes in DASHlink—Flight Data For Tail 687. The maximum altitude within a single flight cycle is over 10000 feet. The flight length ranges from 1 hour to more than 5 hours. Among all the flights, more than half covers a long-length flight of more than 5 hours, while only about 5% of the total flights belong to class no.1. Take unit 14 as an example, the length of the cycle could be far away from three hours.

TABLE I
FLIGHT DATA IN DASHLINK—FLIGHT DATA FOR TAIL 687

| Flight class | Flight Length range(hours) | Number of Flights |
|--------------|----------------------------|-------------------|
| Class no. 1 | 1 to 3 | 18 |
| Class no. 2 | 3 to 5 | 149 |
| Class no. 3 | more than 5 | 185 |

For each unit, data is organized based on flight cycles. As shown in Figure 2, each cycle covers the process of climbing, cruising, and descending. During distinct phases, The input includes the raw data from model health parameters, measurements, and virtual sensors. In order to segment data properly, scenario descriptors are used as the guiding features to decide the modality division. In further studies, the modality division could be modified according to the knowledge of experts.

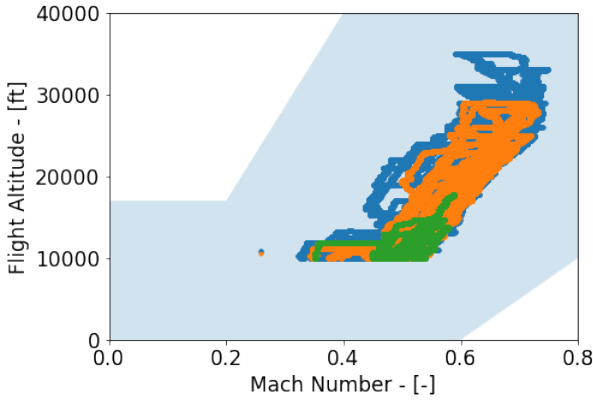


Fig. 2. A overview of the envelope curves of altitude and flight Mach number for Turbofan Engine Degradation Data Set-2. Different colors refer to different flight classes based on flight length range.

To be more precise, scenario descriptors give a brief description of the flight conditions, which consist of the altitude, flight Mach number, Throttle-resolve angle, and the total temperature at the fan inlet. Figure 3 gives an example of these four curves of a single flight cycle. Although the total

length of each cycle varies on a large scale, curves of the scenario descriptors change in a similar pattern. It indicates that the input data could be divided into the same number of modalities according to flight conditions.

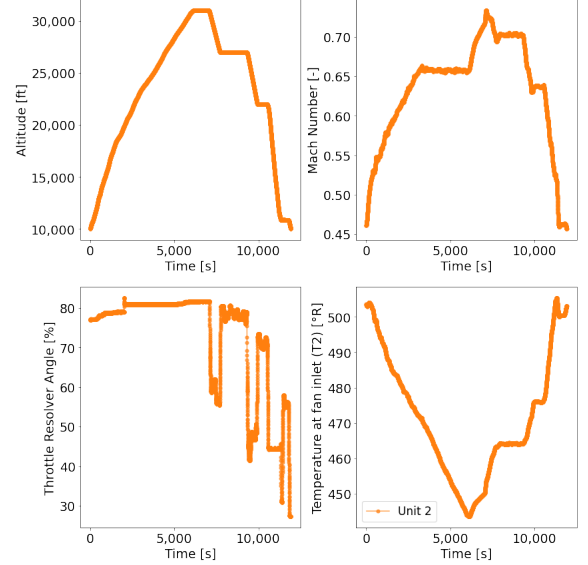


Fig. 3. Altitude, flight Mach number (XM), throttle-resolver angle (TRA) and total temperature at the fan inlet (T2) in a single flight cycle for Unit 5 in DS02.

B. Attention blocks

The Transformer model was first introduced in machine translation, a traditional Natural Language Processing task. Instead of using Convolutional Neural Networks or Recurrent Neural Networks, it only uses multi-head attention to access the features both in encoder and decoder based on the sequence to sequence model. The input sequence is processed with word embedding along with positional embedding, which introduces the positional relationship between elements. Self-attention captures the dependencies between all the elements simultaneously and globally. Consequently, parallel computation becomes possible during the training phase.

1) *Multi-head attention*: When extracting the high-level features from the segmented data, multi-head attention is utilized to acquire the dependencies between different modalities. Multi-head attention provides several heads of self-attention. Each head of the attention module establishes a mapping from the key-value pairs to the output conditioned on the query. What worth mentioning is that the self-attention mechanism is used here. In other words, both the key-value pairs and the queries are computed from the input f , using the following three equations:

$$K_j = fW_j^k \quad (1)$$

$$V_j = fW_j^v \quad (2)$$

$$Q_j = fW_j^q \quad (3)$$

where W_j^k , W_j^v , and W_j^q are trainable weight matrices.

TABLE II
DESCRIPTION OF TURBOFAN ENGINE DEGRADATION DATA SET-2

| Type of descriptors | Number of descriptors | Examples of descriptors |
|-------------------------|-----------------------|--|
| Scenario descriptors | 4 | Altitude, Flight Mach number, TRA, T2 |
| Measurements | 14 | Fuel flow, Physical fan speed, T24, T30, T48 |
| Virtual sensors | 14 | T40, P30, SmFan, Ratio of fuel flow to Ps30 |
| Model health parameters | 10 | Fan efficiency modifier, Fan flow modifier |
| Auxiliary data | 4 | Unit number, Flight cycle number, Flight class, Health state |

The output of each module is the weighted sum of the values. And the weights α evaluate the relevance between the keys and the query. They are computed from the keys and the query using the following equation.

$$y_{att}^j = f_{att}(K_j, Q_j, V_j) = softmax(\frac{Q_j K_j^T}{\sqrt{d_k}}) V_j \quad (4)$$

$f_{att}()$ denotes the self-attention function. A scaled factor d_k is introduced to the dot products. Each head of self-attention is calculated separately, providing a representation. Let h denote the number of heads. Then h results of self-attention are concatenated to form the input of the normalization layer,

$$y_{ma} = Concat(y_{att}^1, y_{att}^2, \dots, y_{att}^h) W^o \quad (5)$$

where W^o is also a parameter matrix.

2) *Feed forward layer*: Before transmitting the features to feed forward layers, residual connection is applied to improve the convergence. The normalized output $LN(x)$ is added with the input of the attention block. The features are then processed by two linear transformation along with a ReLU activation,

$$FFN(x) = ReLU(xW_1 + b_1) \cdot W_2 + b_2 \quad (6)$$

where W_1 , W_2 , b_1 and b_2 are trainable weights of the model.

C. Prediction

The output of the attention blocks is sent to the regression layers for prediction. The high-dimensional features are mapped to another feature space by fully connection. And then a nonlinear function Sigmoid is adopted as the activation function. Consequently, the output of the model is limited to a range between 0 and 1.

$$y_{fc} = y_{att} W_3 + b_3 \quad (7)$$

$$y_{pred} = Sigmoid(y_{fc}) \quad (8)$$

The model uses mean square error as the loss function. The equation is as follows:

$$Loss_{MSE} = \frac{\sum_{i=1}^N (\hat{r}_i - r_i)^2}{N} \quad (9)$$

where r_i denotes the ground-truth label, while \hat{r}_i denotes the predicted RUL of the model.

IV. EXPERIMENT

A. Data Preparation

Turbofan Engine Degradation Data Set-2 presents realistic run-to-failure data organized in units, and about one hundred flight cycles are included in each unit. The total useful life is set to 100 percent. And the label for each flight cycle is calculated with the index of the current cycle divided by the total number of cycles of the unit. In this way, the label is a positive decimal between 0 and 1. The larger the number is, the more numbers of cycles the engine could support.

The data of measurements, virtual sensors, and model health parameters are defined as the input of the model. Due to the variety of flight conditions and the inequality of the flight length during each flight cycle, the data for several certain modalities can be absent. During the training phase, the data set is split in the following way. Seventy percent of the data is used for training, twenty percent for testing, and the rest ten percent for validation.

B. Experimental Setup

The experiments are conducted on a Linux system with 2 Nvidia GTX 1080Ti GPUs. The models are trained for 1000 epochs. For this run-to-failure dataset, the RUL labels are normalized to $[0, 1]$. The details of the network architecture can be found in Table III. The number of attention blocks is set to 6. In each attention block, the number of self-attention heads is set to 16. The hidden dropout ratio is set to 0.1.

TABLE III
DETAILS OF G-TRANSFORMER ARCHITECTURE

| Components | Layer type | Parameter | Value |
|------------------|-----------------------|-------------------------|-------|
| Attention blocks | Multi-head attention | number of heads | 16 |
| | | number of blocks | 6 |
| | | input size | 512 |
| | Layer normalization 1 | input size | 512 |
| | Feed forward | number of hidden neuron | 512 |
| | Layer normalization 2 | input size | 512 |
| Prediction | Fully connection | number of hidden neuron | 128 |
| | Sigmoid | input size | 128 |

Four evaluation metrics, mean absolute error(MAE), root mean square error (RMSE) [16], R2, and scoring function(SF) [17], are utilized to evaluate the performance of G-Transformer model. These four metrics are calculated using the following equations.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{r}_i - r_i| \quad (10)$$

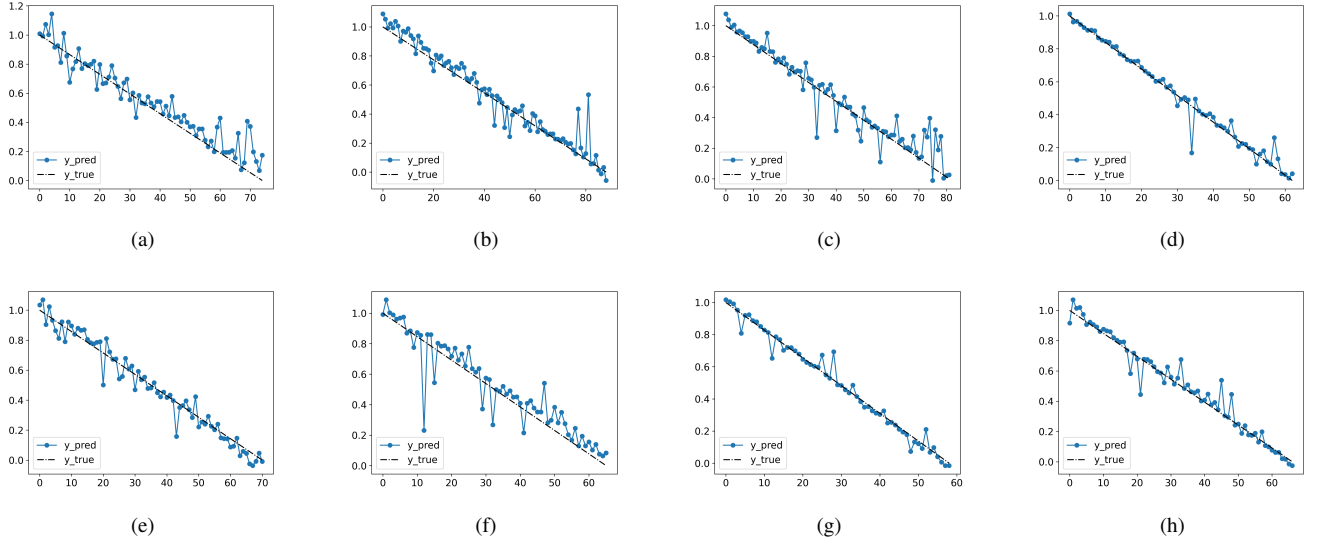


Fig. 4. Curves of predicted results on Turbofan Engine Degradation Data Set-2. Eight subfigures represent the output of the G-Transformer model on units 2, 5, 10, 16, 18, 20, 11, and 15 successively. In each subfigure, the blue curve shows the predicted RULs, while the black dot and dash line displays the desired labels.

TABLE IV
LOSS VALUES AND EVALUATION FUNCTIONS FOR EIGHT UNITS OF N-CMAPSS DATASET.

| Unit | 2 | 5 | 10 | 16 | 18 | 20 | 11 | 15 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|
| Loss | 4.91e-3 | 5.50e-3 | 7.43e-3 | 2.45e-3 | 1.86e-3 | 6.79e-3 | 1.37e-3 | 2.06e-3 |
| RMSE | 7.22e-2 | 7.57e-2 | 8.53e-2 | 4.92e-2 | 4.94e-2 | 9.94e-2 | 3.68e-2 | 5.33e-2 |
| SF | 2.22e-1 | 3.07e-1 | 3.71e-1 | 1.35e-1 | 1.66e-1 | 2.90e-1 | 1.17e-1 | 1.67e-1 |
| MAE | 3.19e-2 | 3.88e-2 | 4.90e-2 | 2.33e-2 | 2.53e-2 | 4.83e-2 | 2.12e-2 | 2.71e-2 |
| R2 | 9.39e-1 | 9.33e-1 | 9.15e-1 | 9.72e-1 | 9.72e-1 | 8.85e-1 | 9.84e-1 | 9.67e-1 |

$$RMSE = \frac{1}{N} \sum_{i=1}^N (\hat{r}_i - r_i)^2 \quad (11)$$

$$R2 = 1 - \frac{\sum_{i=1}^N (\hat{r}_i - r_i)^2}{\sum_{i=1}^N (\hat{r}_i - \bar{r})^2} \quad (12)$$

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N r_i \quad (13)$$

$$SF = \begin{cases} \frac{1}{N} \sum_{i=1}^N (e^{-\frac{\hat{r}_i - r_i}{13}} - 1), & \text{if } \hat{r}_i < r_i \\ \frac{1}{N} \sum_{i=1}^N (e^{\frac{\hat{r}_i - r_i}{10}} - 1), & \text{if } \hat{r}_i \geq r_i \end{cases} \quad (14)$$

C. Result and Analysis

Figure 4 presents the predicted results on Turbofan Engine Degradation Data Set-2. Each column gives the predictions on one of the units. The blue curves represent the predicted remaining useful lives, while the black dot and dash lines represent the ground truth labels. Except for some certain flight cycles, The predicted points are in slight fluctuant condition within a narrow range, which demonstrates the effectiveness of our G-Transformer.

To evaluate the performance quantitatively, four evaluation metrics are implemented here. Mean absolute error, root mean square error (RMSE) [16], R2, and scoring function(SF) [17]

lead to slightly different best models at distinct epochs. As listed in Table IV, R2 values of all the units are very close to 1, which indicates the high similarity of the predictions and ground truth. As for MAE, RMSE, and SF, the smaller these three metrics are, the better the regression performance is. The values of MAE, RMSE, and SF are quite small numbers close to 0, which indicates the high similarity between the predictions and the labels. In contrast to those units with longer flight lengths, units of short length cycles suffer from larger prediction errors. According to the distinguishment of the real flight scenarios, empirical results demonstrate that the length of each cycle affects the modality segmentation directly. Consequently, units with shorter cycles are in lack of adequate information required for prediction, leading to severe sparsity.

To better show the regression performance, visualization of the dimensional-reduced features is displayed. High-dimensional features before the last fully connected layer are extracted and processed by t-SNE [18]. After feature reduction, each sample is represented by a two-dimensional vector. Figure 5 shows the t-SNE results with the data of unit 2 as input. Reduced features of samples are depicted in different colors according to RUL labels. All the data points are located in a contiguous area. Data points of a neighborhood have little difference in RUL labels. And cycles of similar

RUL gain a short distance between each other in this two-dimensional space. In addition, if the RUL varies significantly, there is a large gap between the data points. All of these characteristics contribute to the conclusion that the model has a great ability to address the regression problem in turbofan engine degradation.

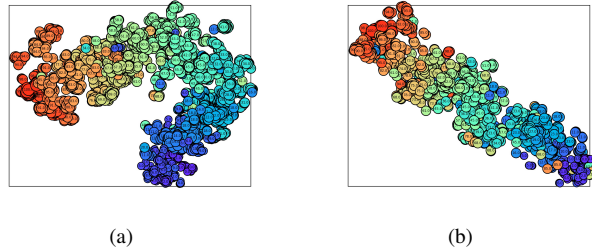


Fig. 5. Visualization of the features before fully connected layer via t-SNE on data of unit 2. The left figure is the visualization of training data, while the right figure shows the results of testing data. The colors of the data points refer to different RUL labels.

V. CONCLUSION

Our G-Transformer model utilizes multi-head self-attention mechanism to capture the features of each modality during flight cycles under real flight conditions. Guiding attributes give access to experts for better modality segmentation. The model is capable of predicting the RUL of turbofan engine degradation under actual flight conditions. Experimental results and analysis demonstrate the effectiveness and robustness of the G-Transformer model with variable-length input. Moreover, the length of flight cycles affects the accuracy directly because of the adequacy of the provided information.

ACKNOWLEDGMENT

The authors acknowledge the supported by the RECLAIM project "Remanufacturing and Refurbishment Large Industrial Equipment" and received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869884.

REFERENCES

- [1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to rul prediction," *Mechanical systems and signal processing*, vol. 104, pp. 799–834, 2018.
- [2] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *2008 international conference on prognostics and health management*. IEEE, 2008, pp. 1–6.
- [3] M. Xia, X. Zheng, M. Imran, and M. Shoaib, "Data-driven prognosis method using hybrid deep recurrent neural network," *Applied Soft Computing*, vol. 93, p. 106351, 2020.
- [4] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.
- [5] H. Li, W. Zhao, Y. Zhang, and E. Zio, "Remaining useful life prediction using multi-scale deep convolutional neural network," *Applied Soft Computing*, vol. 89, p. 106113, 2020.
- [6] H. Yang, K. Ding, R. C. Qiu, and T. Mi, "Remaining useful life prediction based on normalizing flow embedded sequence-to-sequence learning," *IEEE Transactions on Reliability*, 2020.

- [7] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine remaining useful life prediction via an attention-based deep learning approach," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2521–2531, 2020.
- [8] A. Saxena, G. Kai, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, 2008.
- [9] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics," *Data*, vol. 6, no. 1, 2021. [Online]. Available: <https://www.mdpi.com/2306-5729/6/1/5>
- [10] J. Li, X. Li, and D. He, "A directed acyclic graph network combined with cnn and lstm for remaining useful life prediction," *IEEE Access*, vol. 7, pp. 75 464–75 475, 2019.
- [11] T. Xia, Y. Song, Y. Zheng, E. Pan, and L. Xi, "An ensemble framework based on convolutional bi-directional lstm with multiple time windows for remaining useful life estimation," *Computers in Industry*, vol. 115, p. 103182, 2020.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, E. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.
- [14] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.
- [15] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," 2020.
- [16] G. S. Babu, P. Zhao, and X. L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," *Springer, Cham*, 2016.
- [17] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Transactions on Neural Networks and Learning Systems*, "pubMedId": "27416606", 2017.
- [18] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.