WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Next Word Prediction Using Deep Learning

K. Chakradhar[1*],K. Sai Kiran[2],K.Shanmukh[3], K. Sharath Kumar[4],K. Dinesh Sagar[5]

[1,2,3,4,5] GMR Institute of Technology, Rajam, Andhra Pradesh, India

[2] saikirankasi.sk@gmail.com

[1*] chakrkethireedy@gmail.com

ABSTRACT:

Numerous technologies are used to make countless word predicting applications that make typing easier. These technologies that also facilitates typing on a mobile device by suggesting words the end user may wish to insert in a text field. It also increases writing fluency, allowing students to generate more writing skills. It also helps the system for typing of free text. It also helps in forming the sequency of well-structured language. It is used to develop highly recommended applications like Grammarly etc. As well as it is used in user enters that letter of required word, The system displays a list of the most probable words that could appear in the position. It can also predict words various language like Hindi, Spanish etc. The main objective is to predict the next word in a sentence. This study involves N-gram modelling, convolution neural network, recurrent neural networks and some of deep learning techniques which enables the feature information towards predicting the next word more fully. This work also includes results, analysis and techniques. By studying all these we can easily predict the next coming word.

Keywords: Next Word Prediction, N-gram model, Recurrent Neural Network,Long Short-term memory (LSTM) and Natural language processing

Introduction

In this field of Deep Learning and Machine Learning provide various technologies have been developed in today`s world. Every day, we use different devices to type down the textual information and send it to the other end, but it is redundant to type the whole text again and again. So, next word prediction provides user assistance while typing out a sentence and reducing the time consumption during typing. Mostly the rate of transfer of textual information in certain time may increases by predicting the next word. We can also use it for different language. This helps in saving keystrokes of the user. This was extended to continue predicting the next few words for a given sequence of words. Word prediction is an elementary part of Natural Language Generation. It is useful in correcting grammatical error and arrangement of word in a particular process. Word prediction can help early learners like students

or novice researchers to make fewer spelling errors and enhance the speed of typing. Language models assign probabilities to a series of words or a sentence or the probability of the next word given a preceding group of words.It is useful in correcting grammatical error and arrangement of word in a particular process. These models can be useful in a variety of fields, such as spell correction, speech recognition, machine learning. This study involves Natural language Processing, N-gram modelling, Recurrent Neural Network, Deep Learning, Machine learning, Convolutional neural network.

LITERATURE SURVEY

[1] Hamarashid, H. K., Saeed, S. A., & Rashid, T. A. (2021). Next word prediction based on the N-gram model for Kurdish Sorani and Kurmanji. ***Neural Computing and Applications*****,** ***33*****(9), 4547-4566.**

This author includes the main algorithm and their accuracies which performed on predicting Next Word Prediction**.**In this paper, author proposedN-gram language model.Language models assign probabilities to a series of words or a sentence or the probability of the next word given a preceding group of words.In the Stupid Back Off algorithm, the attempt to build language models by distributing true probabilities.These models can be useful in

a variety of fields, such as spell correction, speech recognition, machinelearning, etc.N-gram modelling is utilized to suggest text accurately.The N-gram model has been used for next word prediction to reduce the amount of time.The model is 96.3% accurate.

[2] Yang, J., Wang, H., & Guo, K. (2020). Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network. *IEEE Access*, *8*, 188036-188043.

Multi-window convolution and residual-connected minimal gated unit (MGU) network for the natural language word prediction.The convolution kernels with different sizes are used to extract the local feature information of different graininess between the word sequences.CNN extracts sequence feature information with different granularity by using different window sizes of convolution kernels.The overall experimental results show that the proposed MCNN-ReMGU significantly improves the performance of the word prediction task over the traditional methods.N-gram language model has been used to improve the text input rate in work.It is revealed that 33.36% reduction in typing time and 73.53% reduction in keystroke.The designed system reduced the time of typing free text which might be an approach for EHRs improvement in terms of documentation.

[3] Stremmel, J., & Singh, A. (2021, April). Pretraining federated text models for next word prediction. In *Future of Information and Communication Conference* (pp. 477-488). Springer, Cham.

In this paper, author proposed LSTM-RNN language model for the Next Word Prediction.Federated Averaging Algorithm is used which averages the model parameters. After applying the gradient to all models of database.Federated training on Stack Overflow without any pretraining which yields the two learning curves that exhibit the lowest levels of train and validation accuracy respectively.Dimensionality reduction approach is useful for federated training in which we are constrained by model size.As this paper has achieved the level of accuracy around 22.1%, and 22.2%.Federated learning is a decentralized approach for training models on distributed devices, by summarizing local changes and sending aggregate parameters from local models to the cloud rather than the data itself.

[4] Hamarashid, H. K., Saeed, S. A., & Rashid, T. A. (2022). A comprehensive review and evaluation on text predictive and entertainment systems. *Soft Computing*, 1-22.

This paper has used memory-based learning for next word prediction using Recurrent Neural Network and Long short-term memory.The goal of NLP is to learn and analyze the difficulties of automated generation and comprehending of languages of human beings.RNN has the capability of learning to utilize the previous information when the space between appropriate information and the position that is necessary is tiny. LSTM suffers from a large number of parameters, but it resolves the problem of memory.This model has reached up to the accuracy of 44.2%.Using hybrid-based technique such as Naive Bayes and Latent Semantic Analysis (LSA) model.The probabilistic method Naive Bayes is utilized in NLP like N gram.This model has produced accuracy of 88.2%.

[5] Barman, P. P., & Boruah, A. (2020). A RNN based Approach for next word prediction in Assamese Phonetic Transcription. *Procedia computer science*, *143*, 117-123.

Next word prediction is a highly discussed topic in current domain of Natural Language Processing research.Recurrent Neural Network based language model to improve prediction of next word in sequential data.LSTM to generate complex long-range structured sequences.The Next Word using LSTM with an accuracy of 88.02% for Assamese text and 72.10% for phonetically transcript Assamese language

[6] Rakib, O. F., Akter, S., Khan, M. A., Das, A. K., & Habibullah, K. M. (2019, December). Bangla word prediction and sentence completion using GRU: an extended version of RNN on N-gram language model. In *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE.

Gated Recurrent Unit based Recurrent Neural Network on n-gram dataset.Change in earlier layers will effect output which is shallow. (inefficient).The loss analysis is high using in GRU when compare with Recurrent Neural Network.Average accuracy of 99.70% for 5-gram model, 99.24% for 4-gram model, 95.84% for Tri-gram model, 78.15% and 32.17% respectively for Bi-gram and Uni-gram models on average.

[7] Ambulgekar, S., Malewadikar, S., Garande, R., & Joshi, B. (2021). Next Words Prediction Using Recurrent NeuralNetworks. In *ITM Web of Conferences* (Vol. 40, p. 03034). EDP Sciences.

RNN help to predict next code syntax for users. The used LSTM neural language model to predict within vocabulary words.RNN is reduced no of layers which leads to low accuracy.The short version of LSTM in this CNN try to skip few layers while training result in less training time and they have good accuracy.This model is accuracy of around 56%.

[8] Naulla, N. T. K., & Fernando, T. G. I. (2022, February). Predicting the Next Word of a Sinhala Word Series Using Recurrent Neural Networks. In *2022 2nd International Conference on Advanced Research in Computing (ICARC)* (pp. 13-18). IEEE.

In this research a Recurrent Neural Network has been trained to build a word predictor for Sinhala language. LSTM layers were built and analyzed to obtain the Keras model that displays the highest accuracy and minimum loss.It was observed that more the layers resulted in a decrease of the training and the validation accuracy.It is found that word predictive technologies have several benefits, and one is reducing the keystrokes from 50% to 60% which helps in typing that saves user time and effort.Accurate predictions with a percentage accuracy of 72%.

[9] Shakhovska, K., Dumyn, I., Kryvinska, N., & Kagita, M. K. (2021). An Approach for a Next-Word Prediction for Ukrainian Language. ***Wireless Communications and Mobile Computing, 2021.***

Long short-term memory (LSTM) with a Coupled Input and trained on the server and baseline n-grammodel was compared to the federated learning model trained from scratch. Recurrent NeuralNetworks are a collection of algorithms that is aimed at recognizing patternsThe loss analysis demonstrates that the least lost is in bidirectional RNN.Word prediction is helpful for users because it can boost typing speed and help to omit errors.Accuracy predictions with a percentage accuracy of 75%.

[10] Stremmel, J., & Singh, A. (2021, April). Pretraining federated text models for next word prediction. In ***Future of Information and Communication Conference*** **(pp. 477-488). Springer, Cham.**

Federated Averaging Algorithm is used which averages the model parameters. After applying the gradient to all models of databaseFederated training on Stack Overflow without any pretraining which yields the two learning curves that exhibit the lowest levels of train and validation accuracy respectively.Dimensionality reduction approach is useful for federated training in which we are constrained by model size.As this paper has achieved the level of accuracy around 22.1%, and 22.2%.

[11] Yazdani, A., Safdari, R., Golkar, A., & R Niakan Kalhori, S. (2019). Words prediction based on N-gram model for free-text entry in electronic health records. ***Health information science and systems, 7*****(1), 1-7.**

N-gram is a potential predicting model which measures the probability of the occurrence of a word.Data are not easily analyzed and cannot be linked to the structured records.N-gram language model has been used to improve the text input rate in work.It is revealed that 33.36% reduction in typing time and 73.53% reduction in keystroke.

[12] Aliprandi, C., Carmignani, N., Deha, N., Mancarella, P., & Rubino, M. (2018). Advances in nlp applied to word prediction. ***University of Pisa, Italy February..***

NPL is used large number of word forms makes word prediction for inflected languages a hard task.RNN leads to face vanishing-gradient problem. That is unable to store the long-term dependency between the data.Word prediction is particularly useful to minimise keystrokes for users with special needs, and to reduce misspellings for users having limited language proficiency.A relevant improvement in Keystroke Saving, which now reaches 51%, comparable to what achieved by word prediction methods for non-inflected languages.

[13] Thakur, D. S., Tarsarya, R. N., Vaskar, A. A., & Save, A. A Survey on Text Prediction Techniques Deepal S. Thakur1, Rajiv N. Tarsarya1, Akshay A. Vaskar1, Ashwini Save1.

Proposed RNN for the structure sentence representation.LSTM plays important role, being a memory storage it holds the characters which helps in predicting the next word.It was observed that more the layers resulted in a decrease of the training and the validation accuracy.RNNs are well suited to encode order information and long-range context dependency.As this paper as achieved the level of accuracy around 84.6% and for Recurrent and Convolutional Neural Network around 84.3%.

[14] Rianti, A., Widodo, S., Ayuningtyas, A. D., & Hermawan, F. B. (2022). NEXT WORD PREDICTION USING LSTM. ***Journal of Information Technology and Its Utilization, 5*****(1), 10-13.**

LSTM model is used in this next word prediction.When RNN model was built it has limit the data or the architecture model.It helped us to increase writing fluency and save time.As this paper as achieved the level of accuracy around 75%.

[15]Endalie, D., Haile, G., & Taye, W. (2022). Bi-directional long short term memory-gated recurrent unit model for Amharic next word prediction. ***PloS one, 17*****(8), e0273156.**

In terms of predicting within vocabulary words, a neural language model based on an LSTM network was proposed.BLSTM have been widely used for next word prediction in different languages.The gated recurrent unit performs tasks of natural language processing, speech signal modeling, and music modeling like that of LSTM.The loss analysis shows that bi-directional RNN has the lowest loss and GRU has the highest.GRU works a little faster and bi-directional RNN has the longest execution time.Accuracy of LSTM is 75.02%,GRU is 73.5%,Bi-LSTM-76.1%,Bi-LSTM-78.6%.

METHODOLOGY/RECENT TECHNOLOGY

Stupid Backoff Algorithm:

N-gram modelling techniques are used. And it is divided into sub-techniques like Unigram language model, Bigram language model, Trigram language model.

1.1 Unigram language model: The probability of predicting the present word from set of hole words.

$$P(W_i)=C(W_i)/C(W) \quad (1)$$

1.2 Bigram language model: The probability of predicting a word in combination of previous word from the set of hole words.

$$P(W_i/W_{i-1}) =P(W_iW_{i-1})/P(W_{i-1}) \quad (2)$$

1.3 Trigram language model: The probability of predicting next two in combination of previous of word from the set of hole words.

$$P(W_i/ W_{i-2} ,W_{i-1}) = P(W_{i-2} ,W_{i-1} ,W_i)/P(W_{i-2} ,W_{i-1}) \quad (3)$$

Where $P(W_i)$ is probability of predicting word,

$C(W_i)$ is cost of probability to obtain that word.

$P(W_i/W_{i-1})$ is the probability of predicting next two words,

$P(W_i W_{i-1})$ is probability for obtain next two words,

$P(W_i/ W_{i-2} ,W_{i-1})$ is probability for predicting next three words,

$P(W_{i-2} ,W_{i-1} ,W_i)$ is probability for obtaining next three words.

In the Stupid BackOff algorithm, the attempt to build language models by distributing true probabilities is stopped.If the higher-order N-gram is counted as Zero, then we will backoff to lower-order N-gram.

As a result, a probability distribution is not generated by the Stupid BackOff algorithm.

$$S(W_i/W_{i-k+1}^{i-1})=\{count(W_{i-k+1}^{i})/count(W_{i-k+1}^{i})if$$

$$count(W_{i-k+1}^{i})=\{\lambda S(W_i/W_{i-k+1}^{i-1}) \quad (4)$$

where λ is the weight of BackOff , BackOff variable is made to rely on k.

The Stupid BackOff algorithm is used, which utilizes a few scenarios at the same time. This means discovering probability with a large number of words.For instance, when starting with the trigram and there is not sufficient evidence, then the system backs off to the bigram or unigram.The model is 96.3% accurate.

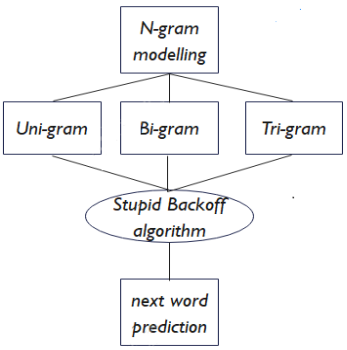

Fig-1: N-gram modelling

Word Embeddings for Federated Training:

In this method includes the implementation of pretraining Word Embeddings.In this process, The post-processing algorithm is used in order to capture different information about word relationships.This post-processing algorithm is followed with Dimensionality Reduction algorithm, Which enhances the experimental runs. (diseried dimensions 100 and 300)This experimental run has achieved FastText.

Algorithm 1: Post-Processing Algorithm PPA(X, D)

Data: Word Embedding Matrix X, Threshold Parameter D

Result: Post-Processed Word Embedding Matrix X

/* Subtract Mean Embedding

1. $X = X - x$

/* Compute PCA Components(principle component analysis)

2.u_i = PCA(X) where i = 1, 2,...,D

/* Remove Top-D Components

3. for all v in X do

4. $v = v - D\ i=1(uT_i \cdot v)u_i$

5. end for

Algorithm 2: Dimensionality Reduction Algorithm PP PCA PP(X, N, D)

Data: Word Embedding Matrix X, New Dimension N, Threshold parameter D

Result: Word Embedding Matrix of Reduced Dimension N: X

/* Apply Algorithm 1 (PPA)*/

1. X = PPA(X, D)

/* Transform X Using PCA to N Dimensions*/

2. X = PCA Transform(X)

/* Apply Algorithm 1 (PPA) */

3. X = PPA(X, D)

Dimensionality reduction approach is useful for federated training in which we are constrained by model size.The same level of accuracy as the paper which achieves 22.1%, and 22.2%.

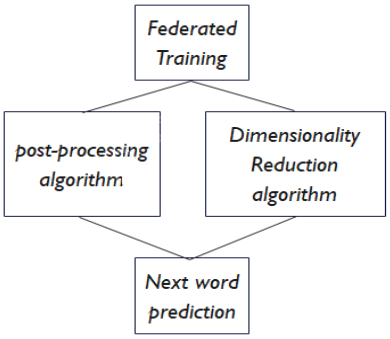

Fig-2: Federated training

BI-DIRECTIONAL LSTM and GRU:

LSTM and Bi-directional LSTM is widely used for next word prediction in different languages. BLSTM learns the bi-directional dependence between terms or words. It results in a more complex network model that takes longer to train.LSTM model learns the sequence of words in a sentence in just one way, either forward or backward.Due to this, the model does not properly understand the sequence.BLSTM-GRU model for predicting the next Amharic word.GRU was utilized as a second layer of BLSTM in the proposed model to minimize network complexity.The gated recurrent unit performs tasks of natural language processing, speech signal modeling, and music modeling like that of LSTM.In GRU, the input and forget gates are combined and controlled by one gate.This combination of gates makes GRU simpler than LSTM.To prevent overfitting, a bidirectional LSTM followed by a dropout layer is used.

Algorithm: Proposed BLSTM-GRU Model Start

1. Input: Amharic sentence sequence
2. Output: next word of the given sequence
3. BLSTM layer (hidden size, batch size)
4. Dropout layer
5. GRU layer (hidden size, batch size)
6. Dropout layer
7. Sigmoid activation function

End

The sigmoid function is that the proposed model predicts probability as an output.This model has produced 78.6% accuracy.

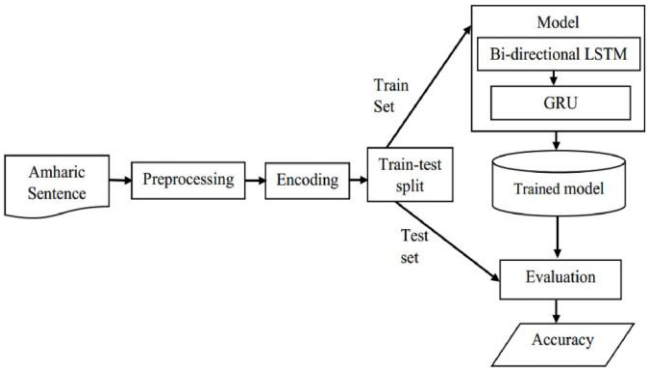

Fig-3: Bi-directional LSTM and GRU

RESULTS & DISCUSSION

In Next word prediction the proposed system utilizes the Stupid Back Off algorithm that, combined with weighted value, builds a list of likely next words. It depends on the N-gram frequencies, the highest frequency for the top five words increases or is matched. Later, it combines the sequences of probability of sequences of words to get the maximum probability of predicting the word. It can be seen that performance increases with the increment of N in the N-gram model. As a result, the straightforward way to assess the used language model is by mapping its accuracy.
The result is the ratio of the correctly predicted words.

Accuracy=Correct prediction/Total prediction

The N-gram model is efficient and easily depends on the frequency of the words instead of complex probabilities that affect the performance of the system, as the system proved to be efficient with the N-gram model.

TABLE-1:N-gram modelling

N-grams	Accuracy (in %)
One-gram	25.4
Two-gram	58.6
Three-gram	72.38
Four-gram	88.24
Five-gram	96.3

While coming to second methodology, It is found that the possibility of reducing the number of federated training rounds required to achieve acceptable model accuracy through the use of pretrained word embedding.Although we use central pretraining with federated fine-tuning which is unable to achieve performance greater than the federated training. It is unable to handle the large data set when compare with federated training. The paper has achieve the level of accuracy around 22.1% and 22.2%.While in third methodology, It is found that the overall estimation that with use of LSTM , GRU, BLSTM model on the Amharic sentence dataset .As the 2.5% more accuracy than RNN model.The proposed predictive network model uses GRU in combination with BLSTM to minimize execution time.

TABLE-2:BILSTM-GRU MODELLING

Algorithm	LSTM	GRU	BLSTM	BLSTM-GRU
Accuracy	75.02%	73.5%	76.1%	78.6%

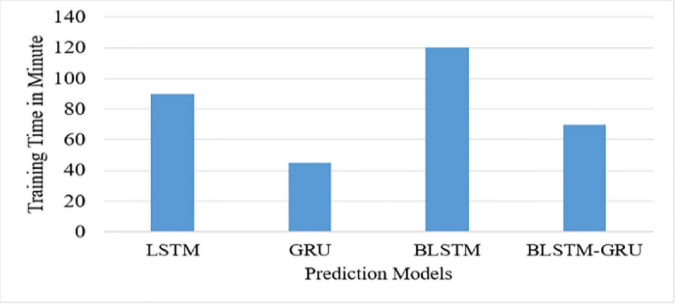

Fig-4:Time graph for BLSTM - GRU

CONCLUSION

As the per the analysis, the first methodology as achieved more accuracy which depends on probability and weights of the given probability in accordance with n-gram modelling. N-gram is a potential predicting model which measures the probability of the occurrence of a word N-gram language model has been used to improve the text input rate in work. The second methodology works long with large data set with less accuracy. It was observed that more the layers resulted in a decrease of the training and the validation accuracy. Federated training on Stack Overflow without any pretraining which yields the two learning curves that exhibit the lowest levels of train and validation accuracy respectively. Natural Language Generation (NLG) is a systematic and significant approach to produce meaningful text that is understandable by humans. Dimensionality reduction approach is useful for federated training in which we are constrained by model size. A Recurrent Neural Network to perform next word prediction on a text sequence and implemented by using a combined approach of RNN and LSTM.The third methodology deal with less running time with best accuracy. LSTM model learns the sequence of words in a sentence in just one way, either forward or backward.As per consideration the first well suitable for next word prediction.

REFERENCES

[1] Hamarashid, H. K., Saeed, S. A., & Rashid, T. A. (2021). Next word prediction based on the N-gram model for Kurdish Sorani and Kurmanji. *Neural Computing and Applications*, *33*(9), 4547-4566.

[2] Stremmel, J., & Singh, A. (2021, April). Pretraining federated text models for next word prediction. In *Future of Information and Communication Conference* (pp. 477-488). Springer, Cham.

[3] Hamarashid, H. K., Saeed, S. A., & Rashid, T. A. (2022). A comprehensive review and evaluation on text predictive and entertainment systems. *Soft Computing*, 1-22.

[4] Barman, P. P., & Boruah, A. (2020). A RNN based Approach for next word prediction in Assamese Phonetic Transcription. *Procedia computer science*, *143*, 117-123.

[5] Rakib, O. F., Akter, S., Khan, M. A., Das, A. K., & Habibullah, K. M. (2019, December). Bangla word prediction and sentence completion using GRU: an extended version of RNN on N-gram language model. In *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE.

[6] Ambulgekar, S., Malewadikar, S., Garande, R., & Joshi, B. (2021). Next Words Prediction Using Recurrent NeuralNetworks. In *ITM Web of Conferences* (Vol. 40, p. 03034). EDP Sciences.

[7] Naulla, N. T. K., & Fernando, T. G. I. (2022, February). Predicting the Next Word of a Sinhala Word Series Using Recurrent Neural Networks. In *2022 2nd International Conference on Advanced Research in Computing (ICARC)* (pp. 13-18). IEEE.

[8] Shakhovska, K., Dumyn, I., Kryvinska, N., & Kagita, M. K. (2021). An Approach for a Next-Word Prediction for Ukrainian Language. *Wireless Communications and Mobile Computing*, *2021*.

[9] Stremmel, J., & Singh, A. (2021, April). Pretraining federated text models for next word prediction. In *Future of Information and Communication Conference* (pp. 477-488). Springer, Cham.

[10] Yazdani, A., Safdari, R., Golkar, A., & R Niakan Kalhori, S. (2019). Words prediction based on N-gram model for free-text entry in electronic health records. *Health information science and systems*, *7*(1), 1-7.

[11] Aliprandi, C., Carmignani, N., Deha, N., Mancarella, P., & Rubino, M. (2018). Advances in nlp applied to word prediction. *University of Pisa, Italy February*..

[12] Thakur, D. S., Tarsarya, R. N., Vaskar, A. A., & Save, A. A Survey on Text Prediction Techniques Deepal S. Thakur1, Rajiv N. Tarsarya1, Akshay A. Vaskar1, Ashwini Save1.

[13] Rianti, A., Widodo, S., Ayuningtyas, A. D., & Hermawan, F. B. (2022). NEXT WORD PREDICTION USING LSTM. *Journal of Information Technology and Its Utilization*, *5*(1), 10-13.

[14] Endalie, D., Haile, G., & Taye, W. (2022). Bi-directional long short-term memory-gated recurrent unit model for Amharic next word prediction. *PloS one*, *17*(8), e0273156.

[15] Yang, J., Wang, H., & Guo, K. (2020). Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network. *IEEE Access*, *8*, 188036-188043.