

Original Research Article

Rolling Bearing Fault Diagnosis based on Residual Neural Network

Haopeng Liang*

College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China. E-mail: 928782706@qq.com

Abstract: Because rolling bearings have been working in an environment with complex and variable working conditions and large noise interference for a long time, the bearing fault diagnosis method has a poor diagnostic effect under variable working conditions. To solve this problem, we propose a residual neural network based on the diagnosis method of rolling bearing fault. The proposed method takes rolling bearing time-domain signal data as input. Because bearing signals have strong time-varying properties, we construct a multi-scale residual block that can not only learn features at different levels, but also expand the width and depth of the residual neural network. We use the advantages of the dilated convolution to expand the receptive field, replace part of the ordinary convolution in the multi-scale residual block with the dilated convolution, and design a multi-scale hollow residual block. The advantage is that the method is made by expanding the receptive field. It has a strong feature learning ability and can learn better features under limited data. Finally, we add a Dropout layer to discard a certain proportion of neurons after the fully connected layer, which can effectively avoid the negative impact of overfitting, and use Case Western Reserve University bearing dataset, the simulation experiment, and the SVM + EMD + Hilbert envelope spectrum, BPNN + EMD + Hilbert envelope spectrum and Resnet three ways of comparative analysis, the results show that the method under the variable condition of the fault diagnosis of rolling bearing has higher diagnosis accuracy, stronger noise resistance, and generalization ability.

Keywords: Rolling Bearing Fault Diagnosis; Residual Neural Network; Variable Condition

1. Introduction

As the core part of rotating machinery, rolling bearing plays a role in supporting the rotating body and reducing friction coefficient during the working process^[1]. As the working conditions of rolling bearings are complex and changeable, various faults may occur, which may lead to accidents and huge property losses^[2]. Therefore, it is very important to diagnose the faults of rolling bearings. In the actual work of rolling bearing, under the action of different loads and under the influence of different noises, the failure of rolling bearing will have a great impact on the performance, stability, and life of

rotating machinery. Therefore, the fault diagnosis of rolling bearing under off-duty working conditions has attracted extensive attention of scholars^[3]. In order to maintain good bearing performance under different working conditions, the rolling bearing fault diagnosis method needs to have a good generalization and anti-noise performance under different load and noise conditions. An effective fault diagnosis method can not only obtain the healthy state of the bearing under off-operating conditions, but also detect the fault type, which is also the most challenging task in rolling bearing

fault diagnosis^[4].

At present, the main methods of rolling bearing fault diagnosis include signal processing method and intelligent diagnosis method, but the signal processing method requires a lot of professional domain knowledge and prior knowledge, which limits its application. Intelligent diagnosis is a new development of mechanical fault detection technology, in which feature-based engineering method and artificial neural network (ANN) based method are two commonly used intelligent diagnosis methods^[5]. The former method mainly realizes fault diagnosis through feature extraction and feature pattern classification. For example, Wang *et al.*^[6] combined the advantages of local mean decomposition and fast peak value to deal with fault characteristics and detected the fault of rotating machinery. MAO *et al.*^[7] used the support vector machine (SVM) as a feature pattern classification algorithm and combined with multi-scale permutation entropy to diagnose the health status of bearings, and achieved a good classification effect. However, the method based on the feature equation still needs to extract feature vectors manually, which has some problems such as strong professional dependence and difficulty in feature extraction, which limits the development of this method in the field of fault diagnosis. With the rise of deep learning, bearing fault diagnosis methods based on artificial neural network (ANN) have been developing rapidly^[8]. For example, Zhang *et al.*^[9] used Fast Fourier Transform (FFT) to transform the original signal as input and achieved good results through a deep convolutional neural network (DCNN) with a wide convolution kernel as the first layer. Chen *et al.*^[10] combined support vector machines and convolutional neural networks to effectively diagnose bearing faults. Jing *et al.*^[11] used a convolutional neural network to learn the characteristics of the time-frequency data of the original data and proved that using a convolutional neural network for feature learning can provide better results than manual feature extraction. Zhang *et al.*^[12] designed a large number of simulation experiments for bearing fault diagnosis, and the results showed that the generalization ability of deep learning method was better than that of the traditional feature extraction fault diagnosis method^[13].

Although deep learning has great potential to avoid false features caused by environmental noise and fluctu-

ations in working conditions, it still requires clever structural design to avoid incorrect features of bearing vibration signals. Therefore, this paper proposes a fault diagnosis method based on residual neural network. The method expands the receptive field by improving the multiscale residual block and adding cavity convolution into the residual neural network, so as to improve the characteristic learning ability of the method. Finally, the Dropout layer is added after the full connection layer, which can effectively avoid the overfitting problem caused by the small amount of effective data in the fault diagnosis of deep neural network, and achieve a better diagnosis effect of rolling bearing.

2. Residual neural network

In 2016, He *et al.*^[14] designed a Residual neural network (Resnet) on the basis of a convolutional neural network, which effectively solved the problem of the disappearance of the gradient of the deep convolutional neural network, and shined in the ImageNet image recognition competition. The neural network introduces the concept of residual learning, which learns residual features through multiple residual blocks connected end to end. The residual block structure is shown in **Figure 1**, where x is the input of the residual neural network, $H(x)$ is the output, $F(x)$ is the residual mapping function, $H(x) = F(x, \{W_i\}) + x$ is the identity mapping function, and Conv is the convolutional layer. The residual neural network adds cross-layers on the basis of ordinary deep convolutional networks, fits the training residuals, and only learns the difference between output and input. He proved by experiment and fitting residual error mapping function than to fitting $F(x) = H(x) - x$ identity mapping function $H(x) = x$ more easily, and in the process of training the underlying error can be quickly connected to the top layer, in addition to using the objective function in the process of making training gradient, also increased the residual gradient, so the residual neural network is While having a deeper number of layers, it also has a stronger feature learning ability.

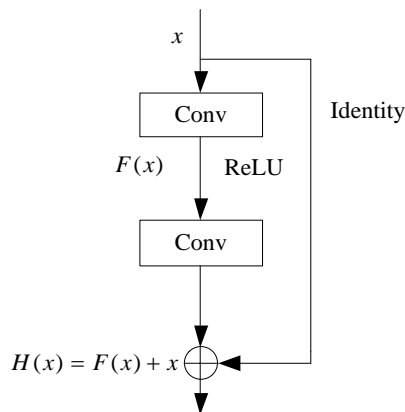


Figure 1. Residual block structure

3. Fault diagnosis method of rolling bearing based on residual neural network

3.1 Multi-scale residual block

Because traditional residual neural network internal residual block contains only one size of convolution kernels, the extracted features are unitary, and often accompanied by changes in load bearing in the operation process, lead to bearing were collected one-dimensional time domain signal has a certain difference, so in order to be kept in the time domain signal extraction to the more abundant characteristic information, this paper constructed a multi-scale residual block, its structure is shown in figure 2.

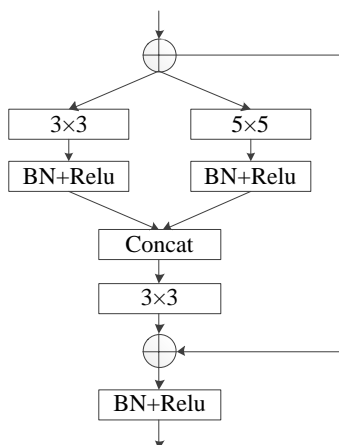


Figure 2. Multi-scale residual block

In Figure 2, the multi-scale residual block improves the first convolutional layer of the traditional residual block. In this layer, two convolution modules of different sizes are used. The size of the convolution kernel is 3×3 and 5×5. Among them, the 3×3 convolution mod-

ule focuses on extracting local features, and the 5×5 convolution module focuses on extracting global features, so that the multi-scale residual block can more effectively extract the feature information in the time domain signal, and in each volume Batch Normalization (BN) and Relu activation function are used in the product modules to accelerate the feature extraction rate of the network. Finally, the two modules are spliced together in the feature dimension through the concat layer and transported to the next convolution layer. We consider that densely connected convolutional layers will increase the amount of parameter calculations and reduce the learning efficiency of the residual neural network, so only use the multi-scale convolution module in the first layer of the residual block. The second layer of the residual block still uses a convolutional layer with a convolution kernel size of 3×3.

3.2 Multi-scale cavity residual block

Because of the difference in bearing signal characteristics under different working conditions, it is difficult to accurately identify the bearing signal characteristic information with the residual neural network constructed by the traditional residual block. To solve this problem, a multi-scale residual cavity block including dilated convolution is designed in this paper. The concept of cavity convolution was proposed by Yu *et al.*^[15] in 2016. Its principle is to introduce a cavity coefficient with an expansion rate d in the ordinary convolution kernel to obtain a new cavity convolution kernel. The width of the cavity convolution kernel is $w + (w - 1)(d - 1)$. The height is $h + (h - 1)(d - 1)$, where w is the width of the ordinary convolution kernel, and h is the height of the ordinary convolution kernel. The convolution operation of the dilated convolution is the same as the ordinary convolution, but the dilated convolution has a larger receptive field and has the same parameter calculation amount as the latter during the network training process. In order to improve the feature learning ability of the method presented in this paper, the ordinary convolution kernel in the multi-scale residual block is replaced by the dilated convolution kernel with an expansion rate of 2 to construct the dilated residual block. In this way, the ordinary convolution kernel has the receptive field of the dilated convolution kernel and has stronger feature learning ability. The structure of multi-scale cavity re-

sidual block is shown in **Figure 3**.

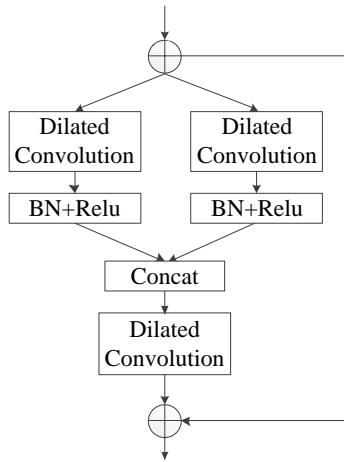


Figure 3. Multi-scale cavity residual block

3.3 Residual neural network method structure

Aiming at the problem of poor fault diagnosis effect caused by the complex working environment of rolling bearings and insufficient effective data samples. We improve the traditional residual neural network. The improved network is mainly composed of 4 residual blocks connected end to end. The structure of this method is shown in **Figure 4**.

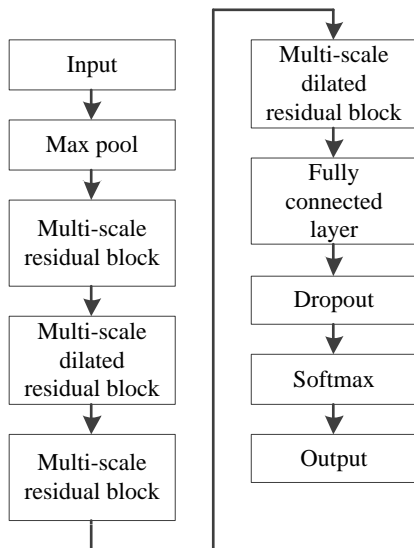


Figure 4. Residual neural network method structure

In **Figure 4**, this method first converts the input signal of the rolling bearing into a two-dimensional grayscale image, and then extracts the characteristic information of the signal through the maximum pooling

layer, and then uses 4 residual blocks. In this paper, the first residual Block and the third residual block are set as multi-scale residual blocks, the second residual block and the fourth residual block are set as multi-scale dilated residual blocks, thereby improving the feature information learning ability of the method, and After each convolutional layer, batch normalization (BN) is added, which improves the generalization ability of the method while increasing the network learning rate. Finally, the extracted data passes through the fully connected layer. The proposed method adds the Dropout layer after the fully connected layer. Dropout is an effective method to alleviate overfitting. During the training process, Dropout will be randomly selected in each training batch. A certain percentage of neurons are discarded, so that the residual neural network only forwards and reverses the parameters of the retained neurons, so as to achieve the effect of regularization and effectively suppress the negative impact of overfitting. Suppose the discard probability of any neuron is P ,

$$p = (p_a = 1 | x) = \sum_{i,j \in B_a} \frac{\exp(w_{i,j}^{(l)} \times x_j^{(l-1)} + b_i^{(l)})}{1 + \exp \sum_{i,j \in B_a} (w_{i,j}^{(l)} \times x_j^{(l-1)} + b_i^{(l)})} \quad (1)$$

Among them, P_a is the probability of sample a generating 1, and B_a is the number of samples a in the i th neuron. In this way, the interaction between neurons and the dependence on some local features can be weakened, so that the method will not be interfered with by overfitting in the fault diagnosis of variable working conditions, and finally output the diagnosis results through the Softmax classifier.

4. Simulation experiment and result analysis

4.1 Experimental data and parameter settings

The experimental data uses the rolling bearing data set published by the Bearing Research Institute of Case Western Reserve University. This data set is an internationally recognized standard data set used to verify bearing failure methods and is used by many scholars. Therefore, using this data set for simulation experiments is also more important and it is very convincing. The

rolling bearing test bench is shown in **Figure 5**.

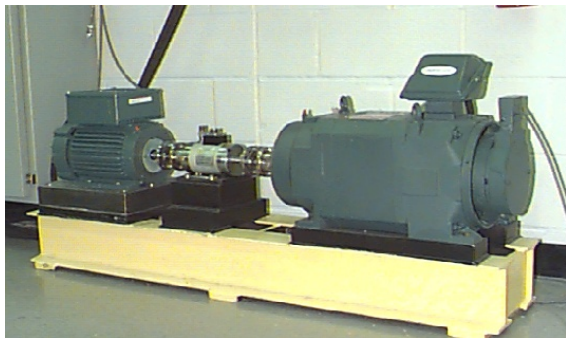


Figure 5. Rolling bearing test bench

In **Figure 5**, the test bench is composed of a motor, an encoder, and a dynamometer, and the test bearing is an SKF6205 motor bearing. The experimental data is collected from the acceleration sensor on the motor drive end and the fan end. There are 4 types of motor speeds, namely 1797, 1772, 1750, and 1730 r/min, corresponding

to load data of 0, 1, 2, and 3hp. The bearing failure positions are at 3 o'clock in the rolling element, inner ring, outer ring, 6 o'clock in the outer ring, and 12 o'clock in the outer ring. The number of sampling channels of the experimental data is 16, and the sampling frequency is 12 kHz. According to the location of the bearing failure and the degree of damage, the collected data is divided into 16 status tags. We set the number of sampling points for each sample to 784, the number of samples in each state label is roughly the same, and the collected data is divided into training samples and test samples at a ratio of 3:1. The test samples under different loads are selected as the variable load test samples, and the variable noise test samples are obtained by loading the Gaussian white noise method with different Signal to Noise Ratio (SNR) on the test samples. The specific experimental data integration table is shown in **Table 1**.

Table 1 Data integration points

Load	Number of training samples	Different load	Number of test samples	
			Variable load	Variable noise
0hp	1972	1hp	740	663
		2hp	740	
		3hp	740	
1hp	2200	0hp	663	740
		2hp	740	
		3hp	740	
2hp	2200	0hp	663	740
		1hp	740	
		3hp	740	
3hp	2205	0hp	663	740
		1hp	740	
		2hp	740	

The simulation experiment platform is TensorFlow, using Python3.7 programming, the computer processor is i5-4200H, the graphics card is NVIDIA GTX860, and the system is Windows10. The experiment uses a small batch training method, set the batch size to 64, the number of iteration batches is 2200, the initial value of the dynamic learning rate is set to 0.001, and the attenuation rate is set to attenuate 0.9 times per 1000 times. In

the training process, the Adam gradient optimization

algorithm is used to update the network parameters, and the dropout is set to a 50% training discard ratio. Dropout is not used in the test process, that is, all neurons are involved in the calculation. In order to prevent the contingency of the experimental results, the average value of 10 experimental results is selected as the final result.

We select the SVM-EMD envelope spectrum,

BPNN-EMD envelope spectrum method, and Resnet method for comparison experiments. The input of the first two methods is the envelope spectrum of the first five eigenmode functions after EMD decomposition, the structure of BPNN is set to 3920-300-16, and the SVM multi-classification method is “one-to-many”, and select “Gaussian kernel function” as its kernel function. The network structure of the Resnet method consists of a data pooling layer, a fully connected layer, and five residual blocks connected end to end. The residual blocks all use a 3×3 convolution kernel.

4.2 Experimental results and analysis

4.2.1 Diagnosis results of variable noise faults

Because the noise generated by the vibration and

mutual friction of the rolling bearing is inevitable during the working process of the rolling bearing, the fault information is difficult to identify, so the rolling bearing fault diagnosis method must have good noise resistance. In order to show that the method in this paper has better advantages in the complex noise environment, it will be compared and analyzed with the three selected fault diagnosis methods. Taking the fault diagnosis of rolling bearing under a load of 2hp and 1750r/min as examples, the signal noise is added to the source test samples. The ratio is 6dB, 9dB, and 12dB Gaussian white noise, and then three variable noise test samples are used to detect the fault diagnosis accuracy of each method. The results are shown in **Table 2**.

Table 2 Accuracy of variable noise fault diagnosis

Different fault diagnosis methods	Accuracy (%)			Average accuracy (%)
	6 dB	9 dB	12 dB	
SVM-EMD	82.86	84.67	83.80	83.77
BPNN-EMD	76.50	76.58	79.30	77.46
Resnet	92.56	96.31	97.85	95.57
Proposed method	95.51	97.70	98.61	97.27

It can be seen from **Table 2** that the fault diagnosis accuracy of the SVM-EMD and BPNN-EMD methods is significantly lower than that of the Resnet method and the proposed method. Although the average fault diagnosis accuracy of the Resnet method is also higher, when the signal-to-noise ratio decreases, the accuracy will decrease. The variable noise fault diagnosis results of the proposed method have higher fault diagnosis accuracy than the other three methods under different SNR, all of which remain above 95%, it can be seen that the method proposed in this paper has better anti-noise ability under the condition of variable noise.

4.2.2 Fault diagnosis results of variable load

When the rolling bearing is working, the load condition will also change, so the generalization ability is often used to measure the variable load diagnosis performance of the rolling bearing diagnosis method. Therefore, this paper conducts experiments on the performance of different methods under variable load conditions. The results are shown in **Figure 6**. **Figure 6** shows the fault diagnosis accuracy of the four methods under

different load changes. The load change refers to the selection of 0hp~3hp load. One of the load data is used as the training set, and the other load data is selected as the test set.

We take the three sets of experimental results of 0-1, 0-2, and 0-3 as examples for analysis. It can be seen from **Figure 6** that the diagnostic accuracy of the two methods, SVM-EMD and BPNN-EMD, are both at below 75%, Resnet’s method can achieve a diagnostic accuracy of up to 92%, while the three groups of experimental diagnosis accuracy based on this method are all above 90%. In contrast, the fault diagnosis accuracy of this method is significantly higher than the previous three methods. In addition, the diagnostic accuracy based on the three methods of SVM-EMD, BPNN-EMD, and Resnet is significantly improved in the two load changes experiments of 2-1 and 2-3. The reason is that the accuracy of the fault diagnosis is from 2hp. When the load changes to 1hp and 3hp, the data representation is relatively similar, so the fault diagnosis accuracy has been improved, but it is still lower than the fault diagnosis accuracy based on

this method. In summary, the method in this paper can better adapt to the difference of rolling bearing vibration signals under variable load and has better generalization ability.

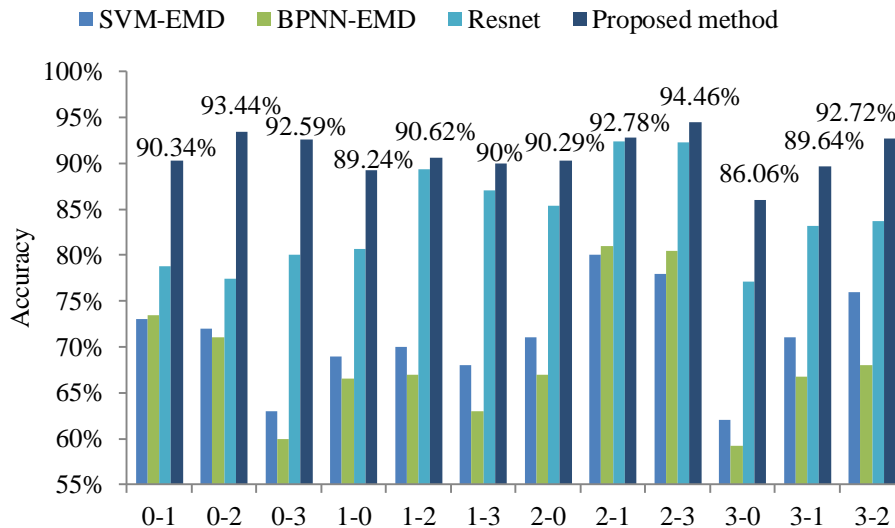


Figure 6. Fault diagnosis results of variable load

5. Conclusion

Aiming at the difficulty of extracting fault features in the time domain signal of rolling bearings and the low accuracy of fault diagnosis, this paper proposes a fault diagnosis method based on residual neural network. The input of this method uses the time-domain vibration signal of rolling bearings. The improved data pooling layer enhances the ability to extract data features, and the designed hollow residual block is added to the residual neural network to improve the efficiency of feature learning. After the fully connected layer enters the Dropout layer to avoid the impact of overfitting and finally output the diagnosis result through the classifier. By comparing with the three methods of SVM-EMD envelope spectrum, BPNN-EMD envelope spectrum, and Resnet, the results show that the residual neural network method proposed in this paper has better anti-noise performance in the experiment of variable noise, can better adapt to the change of load in the experiment of variable load, and has higher generalization ability. Finally, the fault diagnosis accuracy is still higher in the experiment of variable noise and variable load.

About the Author: Haopeng Liang received his B.Sc. degree from Lanzhou University of Technology in

2018. Now he is a master student in Lanzhou University of Technology. His main research interest includes deep learning and fault diagnosis.

References

1. Wang B, Lei Y, Li N, *et al.* Deep separable convolutional network for remaining useful life prediction of machinery. *Mechanical Systems and Signal Processing* 2019; 134.
2. Lu L, Yan J, De Silva CW. Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition. *Journal of Sound and Vibration* 2015; 344.
3. Lu C, Wang Z, Qin W, *et al.* Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing* 2017; 130: 377–388.
4. Zhang X, Liang Y, Zhou J, *et al.* A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* 2015; 69: 164–179.
5. Yan X, Jia M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* 2018; 313: 47–64.
6. Wang L, Liu Z, Miao Q, *et al.* Time-frequency analysis based on ensemble local mean decomposition and fast kurtogram for rotating machinery fault diagnosis. *Mechanical Systems and Signal Pro-*

- cessing 2018; 103: 60–75.
7. Mao W, He L, Yan Y, *et al.* Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mechanical Systems and Signal Processing* 2017; 83: 450–473.
 8. Xiang Z, Zhang X, Zhang W, *et al.* Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO spectrum and stacking auto-encoder. *Measurement* 2019; 138: 162–174.
 9. Zhang W, Peng G, Li C, *et al.* A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* 2017; 17(2): 425.
 10. Chen Z, Gryllias K, Li W. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mechanical Systems and Signal Processing* 2019; 133.
 11. Jing L, Zhao M, Li P, *et al.* A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 2017; 111: 1–10.
 12. Zhang W, Li C, Peng G, *et al.* A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing* 2018; 100: 439–453.
 13. Zhang K, Cheng J, Yang Y. Roller bearing fault diagnosis based on local mean decomposition and morphological fractal dimension. *Journal of Vibration and Shock* 2013; 32(9): 90–94.
 14. He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016.
 15. Yu F, Koltun K. Multi-scale context aggregation by dilated convolutions. *ICLR*; 2016.