

Article

Furniture Image Classification Based on Depthwise Group Over-Parameterized Convolution

Han Ye ¹, Xiaodong Zhu ¹, Chengyang Liu ², Linlin Yang ² and Aili Wang ^{2,*} ¹ College of Materials Science and Engineering, Northeast Forestry University, Harbin 150040, China² Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China

* Correspondence: aili925@hrbust.edu.cn

Abstract: In this paper, an improved VGG16 combined with depthwise group over-parameterized convolution (DGOVGG16) model is proposed to realize automatic furniture image classification. Firstly, depthwise over-parameterized convolution combined with group convolution is combined to construct depthwise group over-parameterized convolution, which is introduced to the VGG 16 model for reducing the number of parameters of the overall model while extracting more sufficient semantic features of furniture images. Then, this paper uses the ReLU activation function in the former part of the neural network to reduce the correlation between parameters and accelerate the weight update speed of the former part of the model. Meantime, the proposed model applies Leaky-ReLU activation function in the last layer to avoid the problem that some neurons do not update. Compared with the six furniture image classification methods based on MobileNetV2, AlexNet, ShuffleNetv2, GoogleNet, VGG 16 and GVGG16, the experimental results show the proposed DGOVGG16 with average accuracy (AA) of 95.51% has better classification performance.

Keywords: furniture image classification; deep learning; VGG16; group convolution; depthwise over-parameterized convolution



Citation: Ye, H.; Zhu, X.; Liu, C.; Yang, L.; Wang, A. Furniture Image Classification Based on Depthwise Group Over-Parameterized Convolution. *Electronics* **2022**, *11*, 3889. <https://doi.org/10.3390/electronics11233889>

Academic Editor: Chunjie Zhang

Received: 3 November 2022

Accepted: 23 November 2022

Published: 24 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of the Internet technology and big data, the number of e-commerce websites has increased dramatically. In the face of such a huge transaction volume, furniture images have become the main information carrier of commodity information transmission. Facing a large number of furniture products, automatic recognition of furniture has become an inevitable requirement for intelligent information extraction in e-commerce [1]. The maturity and development of this technology help businesses automatically identify uploaded furniture images according to color and style. Saving labor consumption can help users quickly search for desired furniture to improve the efficiency of the website and can also help businesses more accurately recommend furniture to improve the sales of furniture. At present, most scholars in the research of furniture image recognition task classify the furniture according to the types of tables, chairs, beds, etc, especially in complex background [2,3].

Traditional image classification methods focus on the extraction of texture, color and other features of the image and select an appropriate classifier to classify the extracted features. With the development of big data technology and the computer hardware performance, deep learning (DL) has become a research hotspot of computer theory [4]. The image classification methods based on convolutional neural network (CNN) are deeply loved by industry and academia [5,6].

The realization of image classification technology is divided into traditional image classification methods and those based on CNN. The two types of recognition methods are composed of two stages: image processing and recognition classification. In the image processing stage, the image to be recognized is input into the system, and the

image is preprocessed appropriately to improve the image quality. Traditional methods usually use special feature extraction algorithms to extract image features and reduce the dimensions of the features to obtain digital features that can accurately describe the image. In the recognition and classification stage, the corresponding statistical learning and other classification decision-making methods are mainly used to match, identify and classify the extracted features. Finally, the classification results are output.

At present, people have tried to use DL to classify furniture images. The convolution machine network does not require complicated feature extraction operations. It uses its own convolution layer and pooling layer to extract features and reduce dimensions, and uses the full connection layer as a classifier to complete image recognition tasks. Zhu Bin [7] et al. studied the emotional intention recognition of chairs based on convolutional neural network technology and built a product image recognition model. Wang Y et al. [8] proposed a metric learning algorithm combined with CNN to remove duplicate and irrelevant samples in the furniture image database. Zhu Haipeng [9] et al. took the chair as an example, and summarized four furniture styles of the East, the West, modern and new technology through expert evaluation and cluster analysis. They used the classic ResNet50 model to recognize the chair image and obtained good recognition accuracy.

In this paper, we combine the advantages of group convolution and depthwise over-parameterized convolution (DO-Conv), proposed depthwise group over-parameterized convolution, which is introduced into VGG16 to improve the classification performance of furniture images. In order to fully mine the semantic information of furniture under complex background, we have carried out many experiments and analyses to prove which layer of VGG16 can get the best classification accuracy. Section 2 introduces the related work of group convolution and depthwise over-parameterized convolution and Section 3 gives the furniture image classification model based on DGOVGG16. In Section 4, the proposed DGOVGG16 model is validated compared to six other classification models. Finally, we gave the conclusions at the end of this paper.

2. The Related Work

2.1. Group Convolution

The group convolution (GC) first appeared in AlexNet, further optimized in ResNeXt network, and solved the problem of insufficient video memory [10]. The first advantage of group convolution is efficient training. Because the convolution is divided into multiple paths, each path can be processed separately by different GPUs, so the model can be trained on multiple GPUs in parallel. Compared with completing all tasks on a single GPU, this model parallelization on multiple GPUs enables the network to process more images at each step. It is generally believed that model parallelization is better than data parallelization. The latter is to divide the dataset into multiple batches, and then train each batch separately. However, when the batch size becomes too small, we essentially perform random gradient descent, not batch gradient descent. This results in slower and sometimes worse convergence results. When training very deep neural networks, packet convolution is very important, as in ResNeXt.

Meantime, with the same convolution operation, GC requires fewer parameters than ordinary convolution to avoid overfit [11]. The difference between ordinary convolution (upper part) and GC (lower part) is as shown in Figure 1, which means the model parameters will decrease as the number of filter groups increases.

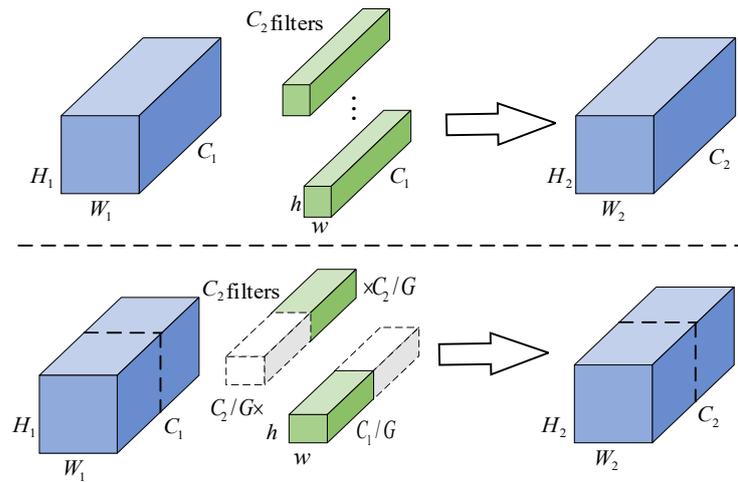


Figure 1. The structure of group convolution.

The GC first groups the input data, and then convolves each group of data separately. The input feature map size of C_1 are divided into G groups as follows:

$$\begin{bmatrix} H_1(x) \\ H_2(x) \\ \vdots \\ H_G(x) \end{bmatrix} = \begin{bmatrix} W_{11}^P & 0 & 0 & 0 \\ 0 & W_{11}^P & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W_{GG}^P \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_G \end{bmatrix} \tag{1}$$

$$F(x) = \sum_{i=1}^G H_i(x) \tag{2}$$

where, $[x_1, x_2, \dots, x_G]$ is the input feature map which is divided into groups size of G , each group has G_1/G feature map size of $H_1 \times W_1$; W_{ii}^P is the convolution kernel of each group with the number of $P = C_2/G$ and size of $h \times w$; $H_i(x)$ is the output feature maps of each group size of $H_2 \times W_2$. Then, all the output feature maps are merged by channel to get the final output of the convolution layer.

If $C_1 = C_2 = C$ which means the input and output feature maps of the above GC are equal, T_{GC} is the number of parameters, which can be calculated as follows:

$$T_{GC} = G \times \left(\frac{C}{G} \times h \times w \times \frac{C}{G}\right) = \frac{h \times w \times C^2}{G} \tag{3}$$

In the same case, the number of parameters of ordinary convolution named as T_{RC} can be calculated as follows:

$$T_{RC} = C \times h \times w \times C = h \times w \times C^2 \tag{4}$$

By comparison of T_{GC} and T_{RC} , the parameter amount of the group convolution is $1/G$ of the standard convolution, which means G is larger, the complexity is lower and the parameter efficiency is higher.

2.2. Depthwise Over-Parameterized Convolution

Li et al. proposed DO-Conv to speed up the convergence of the model, while adding learnable parameters to the model without increasing the computational complexity [12]. DO-Conv is a combination of an ordinary convolution $W \in R^{C_{out} \times D_{mul} \times C_{in}}$ and a deep convolution $D \in R^{(M \times N) \times D_{mul} \times C_{in}}$. Figure 2 shows the flowchart of DO-Conv. P is the size of the input feature map. M and N are the spatial dimensions of P , C_{in} is the number of input feature maps, D_{mul} is the number of deep convolutions, C_{out} is the number of output feature maps, $D^T \in R^{D_{mul} \times (M \times N) \times C_{in}}$ is the transposition of $D \in R^{(M \times N) \times D_{mul} \times C_{in}}$,

and W' is the convolution kernel of. First, combine the deep convolution kernel D^T and conventional convolution kernel W to form W' , and then output O by convolution of W' and P , as shown in Formula (5).

$$O = (D, W) \circledast P = (D^T \circ W) * P \tag{5}$$

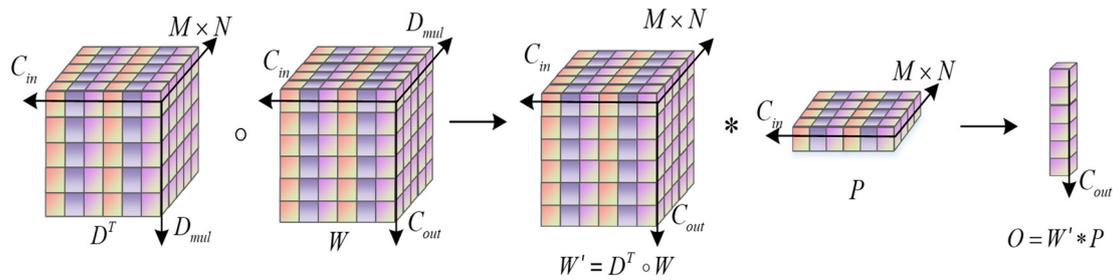


Figure 2. The flowchart of DO-Conv.

2.3. Activation Function

In the DGOVGG16 model, in order to ensure that the weight of the model is updated quickly, alleviate the problem of over fitting, and at the same time avoid losing part of the characteristic information in the negative signal, the ReLU and Leaky ReLU activation function are used in the model to activate neurons. The ReLU activation function is introduced in the previous part of the neural network, so that the output of some neurons is 0, reducing the correlation between parameters and accelerating the weight’s update speed of the previous part of the model. The ReLU activation function written is as follows:

$$\text{ReLU} = \begin{cases} x_i & x_i > 0 \\ 0 & x_i \leq 0 \end{cases} \tag{6}$$

However, if all of the ReLU activation functions are replaced with Leaky ReLU activation functions, the updating speed of the whole model will be reduced. Therefore, this paper only introduces a Leaky value into the last layer of activation functions. The formula of Leaky ReLU function is as follows:

$$\text{Leaky-ReLU} = \begin{cases} x_i & x_i > 0 \\ a_i x_i & x_i \leq 0 \end{cases} \tag{7}$$

where $a_i = 0.01$ is a very small weight, that is, the negative input information is retained without increasing the calculation cost of model training, and the input information is retained more completely.

3. Proposed Methods

The improved VGG16 based on depthwise group over-parameterized convolution (DGOVGG16) model for furniture classification is shown in Figure 3.

First, the furniture data are put into the DGOVGG16 model, and the first 11 layers of the model are used to extract the deep features of the furniture data. The last two layers are divided into depthwise group over-parameterized convolution and conventional convolution layers, respectively. Through depthwise group over-parameterized convolution, more precise features are captured while reducing the number of parameters to ensure the accuracy of the network model to the maximum extent. Secondly, the output features are mapped to the full connection layer through the conventional convolution layer, and after each full connection layer, the dropout layer is introduced, which temporarily discards some neurons from the network at a certain rate, thus eliminating or weakening the con-

nection between neuron nodes, reducing the phenomenon of over fitting, and outputting the final category through the last full connection layer.

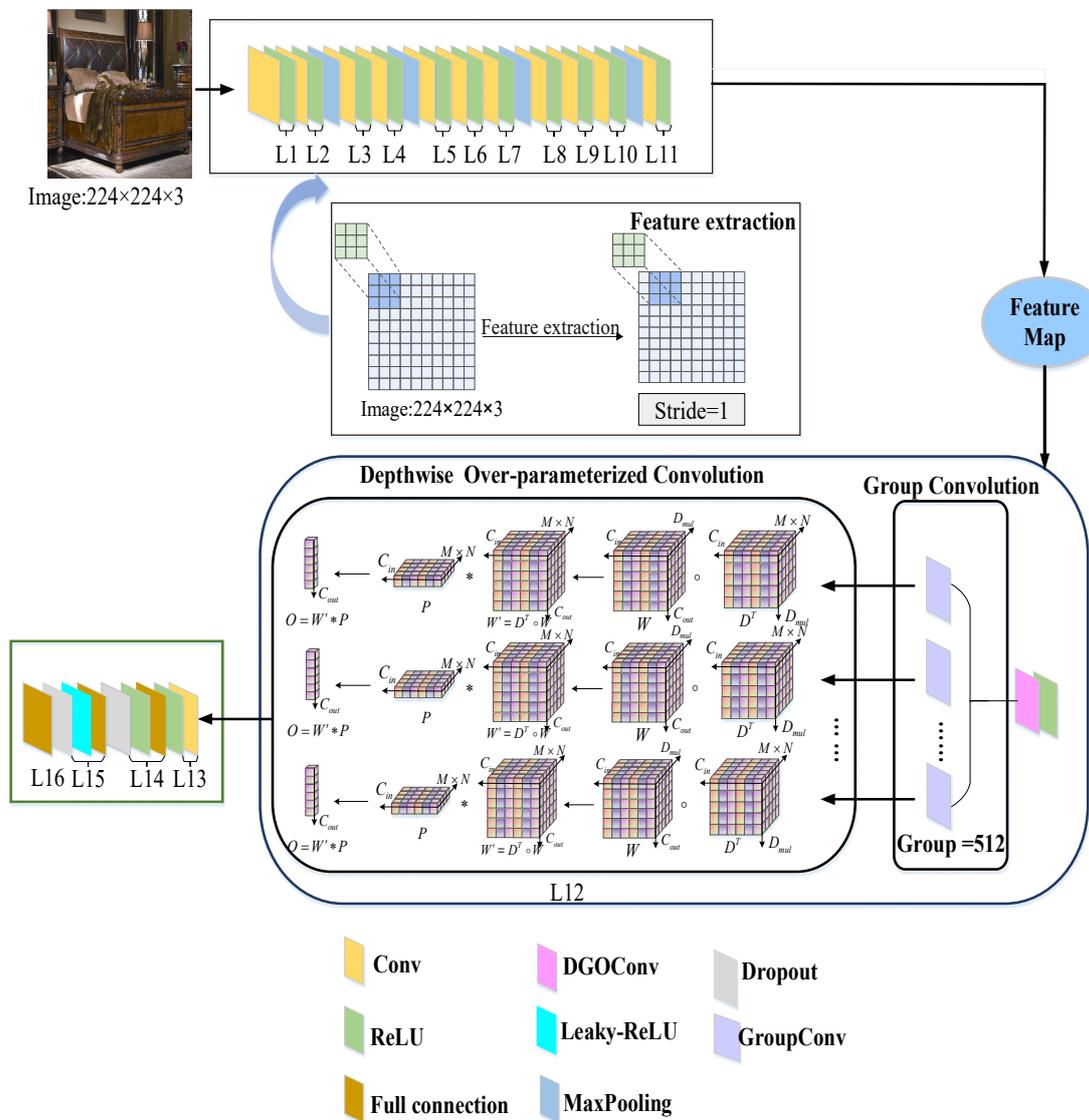


Figure 3. The proposed DGOVGG16 model of furniture image classification.

3.1. DGOVGG16 Model

For the selection of convolutional neural network model, the recognition accuracy and the training convergence efficiency of the model must be comprehensively considered. Generally speaking, the deeper the network model is, and the larger the scale is, the higher the accuracy of image recognition will be. However, the more the corresponding model parameters are, the lower the recognition efficiency will be. Especially when the data set is insufficient, not only the risk of over fitting exists, but also the difficulty of model training will increase difficulty in convergence and other problems. The architecture of the proposed DGOVGG16 network is shown in Figure 4a, and the convolution processing is shown in Figure 4b.

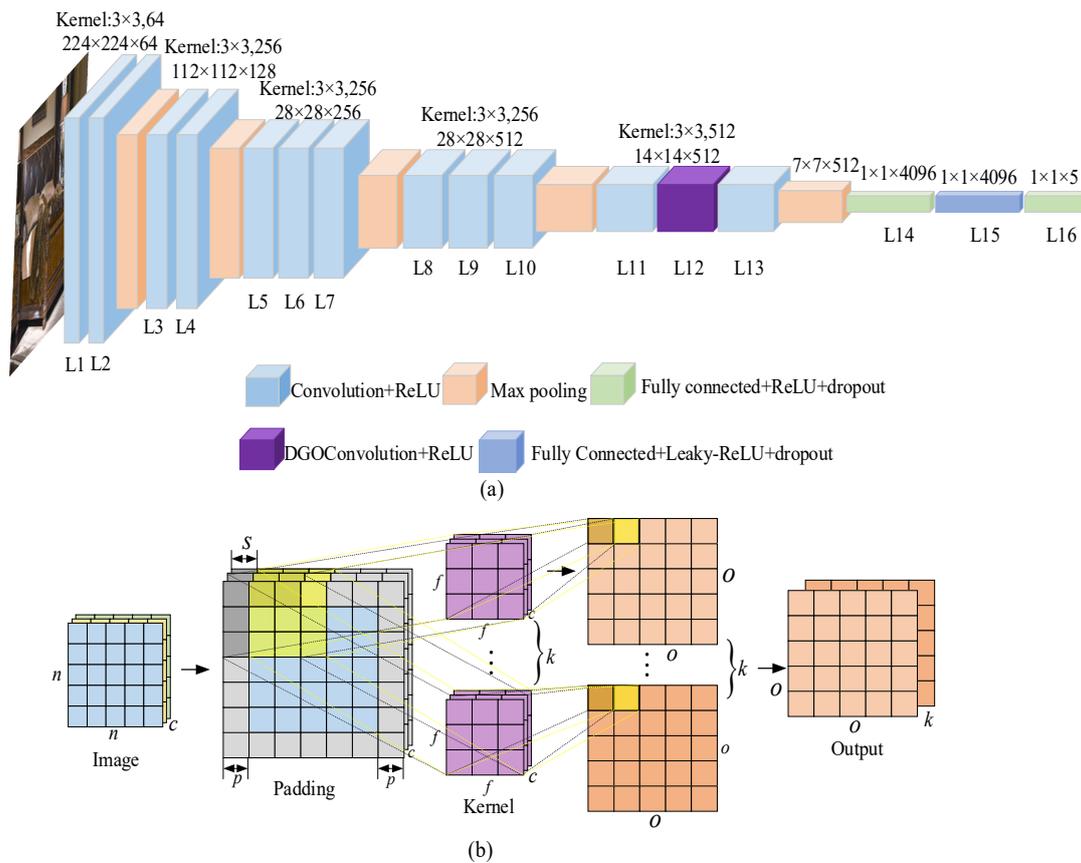


Figure 4. The designed DGOVGG16 model details of each convolutional layer (a) DGOVGG16 structure. (b) convolution processing.

As there is no large-scale image public data set for furniture data set, the accuracy and efficiency of actual furniture classification are also considered. Therefore, this paper chooses to improve the VGG16 network model, which has the advantages of smaller convolution kernel, fewer parameters and strong nonlinear fitting ability. DGOVGG16 model stacks 3×3 small size convolution kernel to replace large convolution kernel for achieving the purpose of using fewer parameters to extract more features, and 2×2 to retain more image information. The number of convolution kernel in the 15 convolution layers is 64, 64, 128, 128, 256, 256, 256, 512, 512, 512, 512, 512, 512, 512, 512, respectively. The number of neurons in the two fully connected layers are 4096, and the number of neurons in the last output layer is 5. The pooling layer adopts the maximum pooling method, and the two full connection layers adopt ReLU and Leaky ReLU activation function respectively. Among them, the ReLU function is easy to derive and fast to calculate, and make the output of some neurons zero, thus reducing the correlation between parameters, which can effectively alleviate the occurrence of overfitting problems. At the same time, the introduction of Leaky ReLU activation function in the last layer can effectively prevent some neurons from being invalid when the final input is negative and ensure the final accurate output results.

At the same time, in order to further prevent overfitting, a dropout layer is added to the DGOVGG16 model. Its main idea is to temporarily discard some neurons from the network according to a certain ratio, so as to eliminate or weaken the connection between neural nodes. The dropout process reduces the size of the parameters involved in training. If the total number of neural nodes in the convolutional network structure is n , the combination size of 2^n in theory can be constructed. Such a combination can not only increase the size of the parameters, but also improve the generalization of the network. Moreover, dropout has a great probability to create different combinations of neurons working together each time.

The combination of these different neurons reduces the node correlation, thus enhancing the generalization ability of the entire model.

3.2. Depthwise Group Over-Parameterized Convolution

Although VGG16 model itself has the advantages of smaller convolution kernel and fewer parameters, there is still a risk of over fitting with the deepening of network layers.

Therefore, in the 12th layer of VGG16 model, this paper introduces depthwise group over-parameterized convolution as shown in Figure 3, which is composed of two parts, namely, group convolution and DO-Conv. Among them, the number of parameters of the network model is minimized by grouping to avoid overfitting. The reduction of the number of parameters means that the complexity of the model is reduced. The lower the complexity, the lower the accuracy of the model will also decrease. Therefore, this paper conducts depthwise over-parameterized convolution for each input channel, which can fully extract the semantic features of the furniture image ensuring that the overall model to extract more sufficient semantic features while reducing the amount of parameters.

4. Experimental Results and Analysis

4.1. Furniture Dataset Description

The furniture dataset used in this paper is the open data set of Kaggle competition, which was released by Nikhil Akki at Deloitte, a data science engineer from Mumbai, India, in 2018. Some images of furniture data are shown in Figure 5, which can be divided into five categories: bed, chair, sofa, swivelchair, and table. It can be seen from Figure 5 that the background of furniture is usually complex, including invalid information unrelated to furniture such as walls, windows, curtains, external environment, people, etc. Among them, there also exists that chair appears in the table sample as shown the last column. Therefore, when classifying furniture images, it is necessary to take into account the complexity of its background and the diversity of furniture features and design an accurate classification model of furniture in a complex background. The specific number of each type of furniture is shown in Table 1.

Table 1. The description of furniture dataset.

Sample Image	Class	Number
	bed	1100
	chair	1100
	sofa	1100
	swivelchair	1100
	table	470



Figure 5. The samples of the furniture dataset. (a) bed. (b) chair. (c) sofa. (d) swivelchair. (e) table.

4.2. Experimental Platform Parameter Settings

The experimental environment was Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz processor, 128 GB running memory (RAM), NVIDIA GeForce RTX 2080Ti GPU. In addition, the deep learning framework was Pytorch. In the experiments, the size of the input data is set to 224, the number of training epochs is set to 30, and SGD optimization algorithm is used for optimization. The momentum is set to 0.9, and the learning rate is set to 0.1. All experimental training samples are randomly selected from furniture dataset and the ratio of training set to test set is 10:1. Meanwhile, this paper uses six different methods for comparison and analysis, including MobileNetV2, AlexNet, ShuffleNetv2, GoogLeNet [13], VGG16 and VGG16 network based on group convolution (GVGG16). The average accuracy (AA) commonly is used in furniture image classification as the evaluation index to better evaluate the proposed DGOVGG16 model's classification performance of each category.

4.3. Design of DGOVGG16 Model

This paper takes DGOVGG16 as the main feature extraction network, which consists of 15 ReLU layers, 12 convolution layers, 5 MaxPool layers, 3 full connection layers, 1 over-parameterized convolution layer, 1 Leaky-ReLU layer, and 2 dropout layers. In order to make DGOVGG16 model use fewer parameters to represent more features, this paper sets the size of convolution kernel as 3×3 to replace the large convolution kernel by stacking small convolution kernels, and the size of the largest pooling layer is 2×2 to retain more image information. The number of neurons in the three full connection layers was 4096, 4096 and 5, respectively. When the dropout rate is 0.5, the experimental accuracy is the highest. Table 2 shows the parameter settings of DGOVGG16 network model.

Table 2. The parameter settings of DGOVGG16 model.

Layer	Kernel	MaxPool	ReLU	Dropout	Leaky-ReLU
Conv1	3 × 3 (stride = 1)	No	Yes	No	No
Conv2	3 × 3 (stride = 1)	2 × 2 (stride = 2)	Yes	No	No
Conv3	3 × 3 (stride = 1)	No	Yes	No	No
Conv4	3 × 3 (stride = 1)	2 × 2 (stride = 2)	Yes	No	No
Conv5	3 × 3 (stride = 1)	No	Yes	No	No
Conv6	3 × 3 (stride = 1)	No	Yes	No	No
Conv7	3 × 3 (stride = 1)	2 × 2 (stride = 2)	Yes	No	No
Conv8	3 × 3 (stride = 1)	No	Yes	No	No
Conv9	3 × 3 (stride = 1)	No	Yes	No	No
Conv10	3 × 3 (stride = 1)	2 × 2 (stride = 2)	Yes	No	No
Conv11	3 × 3 (stride = 1)	No	Yes	No	No
DOConv12	3 × 3 (stride = 1) (groups = 512)	No	Yes	No	No
Conv13	3 × 3 (stride = 1)	2 × 2 (stride = 2)	Yes	No	No
FC	Neurons = 4096	No	Yes	Yes	No
FC	Neurons = 4096	No	No	Yes	Yes
FC	Neurons = 5	No	No	No	No

4.4. The Selection of Depthwise Group Over-Parameterized Convolutional Layer

The group convolution in depthwise group over-parameterized convolution applied over-parameterized convolution to the input grouped channels. In order to minimize the parameter amount of the proposed model, this paper sets the number of groups divided as the number of input channels. From the number of channels input at each layer, it can be found that layers 2 to 3 are equal, layers 4 to 5 are equal, layers 6 to 8 are equal, and layers 9 to 13 are equal. Moreover, the features extracted from the next layer are more accurate than those extracted from the previous layer. Thus, this paper introduces depthwise group over-parameterized convolution into layer 3, layer 5, layer 8 and layer 12 respectively for comparison, and selects the most appropriate number of layers. It can be seen from Table 3 that when the depthwise group over-parameterized convolution is put into the 12th layer, the AA growth reaches the peak, for 12th layer extracts more refined furniture semantic features with stronger discrimination. Thus, this paper sets the location of depthwise group over-parameterized convolution as the 12th layer.

Table 3. The classification accuracy comparison of depthwise group over-parameterized convolution in different layer.

Classes	Layer	3	5	8	12
	bed		81.00	96.00	83.00
chair		91.00	87.00	90.00	98.00
sofa		84.00	93.00	94.00	95.00
swivelchair		85.00	88.00	88.00	94.00
table		91.30	82.61	86.96	94.56
AA(%)		86.46	89.32	88.40	95.51

4.5. Comparison of Experimental Results

This paper makes a comparative analysis of the two evaluation indicators: AA and running time. Table 4 shows that DGOVGG16 model is superior to other six classification

models. The comparison between DGOVGG16 and GVGG16 shows that DGOVGG16 performs better than the latter. The DO-Conv is introduced to reduce the number of model parameters while fully extracting the semantic features of furniture, reducing the problem of model overfitting. At the same time, Leaky-ReLU activation function is introduced to retain the update weight speed of the whole model. Therefore, the DGOVGG16 model proposed in this paper has a stronger classification accuracy compared with other classification methods.

Table 4. The comparison of furniture classification results based on different network models.

Classes	MobileNetV2	AlexNet	ShuffleNetv2	GoogLeNet	VGG16	GVGG16	DGOVGG16
bed	94.40	88.60	94.00	95.40	93.40	91.00	96.00
chair	94.60	95.20	97.00	95.80	96.40	93.00	98.00
sofa	88.00	88.60	92.00	95.20	92.20	92.00	95.00
swivelchair	89.60	91.00	89.00	88.00	95.00	93.00	94.00
table	88.04	91.30	91.30	89.56	95.65	91.30	94.56
AA (%)	90.92	90.94	92.66	92.79	94.53	92.06	95.51
Time(s)	729	675	688	722	769	746	747

It can be seen from Table 4 that in the furniture dataset classification experiment of DGOVGG16, AA reached 95.51% at the highest, and increased by 3.45%, 0.98%, 2.72%, 2.85%, 4.57% and 4.59% respectively compared with GVGG16, VGG16, GoogLeNet, ShuffleNetv2, AlexNet and MobileNetV2. AlexNet model is the classification method with the shortest running time. It can be found that the running time of DGOVGG16 is 72 s longer than that of AlexNet, but the accuracy is increased by 4.57%, thus a slight increase in time is acceptable. Figures 6 and 7 show the classification results of different methods and each category on the furniture dataset respectively.

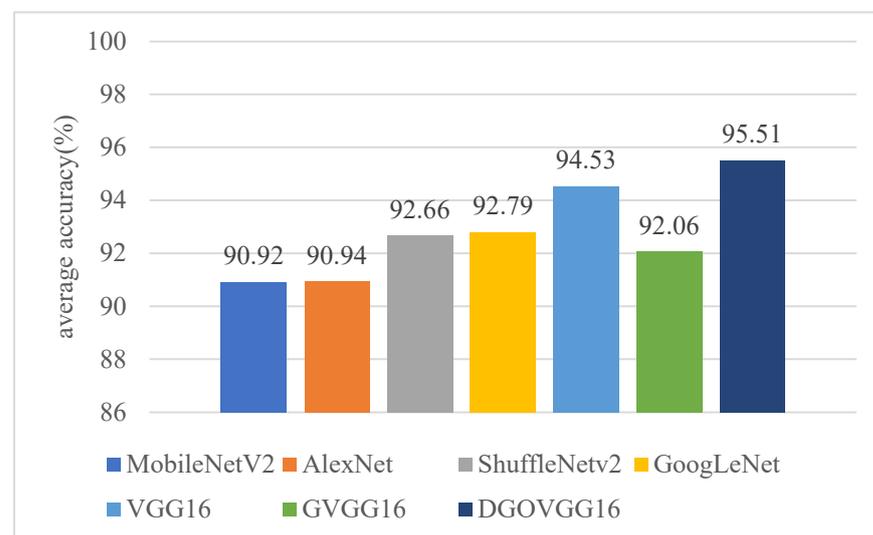


Figure 6. Classification accuracy comparison for different models on the furniture dataset.

It can be seen from Figure 7 that the accuracy of each category of DGOVGG16 model has been significantly improved compared with that of GVGG16 and VGG16, GoogLeNet, AlexNet, MobileNetV2 and ShuffleNetv2 methods. Although some categories are not of the highest accuracy, the overall performance of DGOVGG16 is the best.

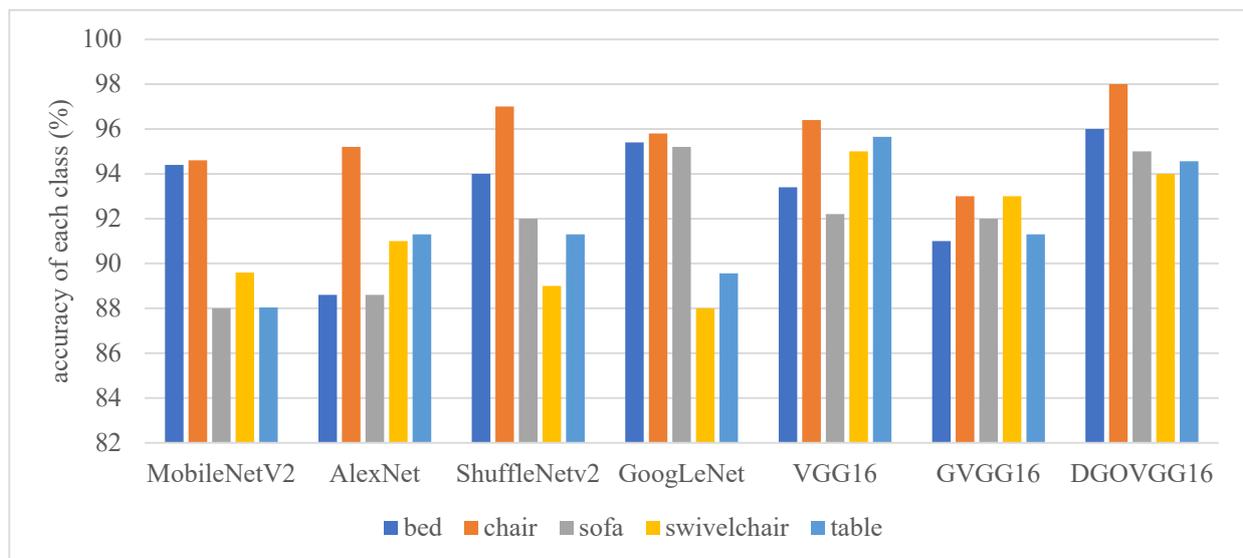


Figure 7. Classification accuracy of each class in the furniture dataset.

5. Conclusions

This paper proposes a furniture classification algorithm based on VGG16 network combined with depthwise group over-parameterized convolution. In VGG16 model, depthwise group over-parameterized convolution is introduced to reduce the number of parameters of the overall model while extracting more sufficient semantic features, which reduces the risk of model overfitting and improves the furniture image classification accuracy of the model, especially in complex backgrounds situation. In addition, in the selection of activation function, this paper used ReLU activation function in the former part of the neural network, reduced the correlation between parameters, accelerated the weight update speed of the former part of the model, and applied Leaky ReLU activation function in the last layer to avoid the problem that some neurons do not update.

This paper conducts experiments on the furniture public data set and compares DGOVGG16 with other six popular classification network methods. The AA of the proposed DGOVGG16 in this paper reaches 95.51%, which is superior to other classical methods, which fully shows that our model can accurately classify furniture images under complex backgrounds with stronger competitiveness. In future work, we intend to mine and fuse the global and local features of furniture at different scales improve the proposed DGOVGG16 model, so that the classification accuracy between similar furniture images can be further improved.

Author Contributions: Conceptualization, H.Y. and X.Z.; methodology, H.Y., X.Z. and A.W.; software, H.Y., C.L. and L.Y.; validation, X.Z., H.Y., C.L. and L.Y.; writing—review and editing, X.Z. and A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by “Belt and Road Initiative” Innovative Talent Exchange Foreign Experts Project (G2022012010).

Data Availability Statement: <https://www.kaggle.com/datasets/akkithetechie/furniture-detector/download?datasetVersionNumber=1> (accessed on 2 November 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
- Mo, K.; Zhu, S.; Chang, A.X.; Yi, L.; Tripathi, S.; Guibas, L.J.; Su, H. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.

3. Varvadoukas, T.; Giannakidou, E.; Gomez, J.V.; Mavridis, N. Indoor Furniture and Room Recognition for a Robot Using Internet-Derived Models and Object Context. In Proceedings of the 10th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 17–18 December 2012.
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
5. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
6. Xiao, J.; Hays, J.; Ehinger, K.A.; Oliva, A.; Torralba, A. SUN database: Large-scale scene recognition from abbey to zoo. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3485–3492.
7. Zhu, B.; Yang, C.; Yu, C.; An, F. Product Image Recognition Based on Deep Learning. *J. Comput. -Aided Des. Comput. Graph.* **2018**, *30*, 1778–1784. [[CrossRef](#)]
8. Wang, Y.; Gao, W.; Wang, Y. Application of furniture images selection based on neural network. In Proceedings of the 6th International Conference on Computer-Aided Design, Manufacturing, Modeling and Simulation (CDMMS 2018), Busan, Republic of Korea, 14–15 April 2018.
9. Zhu, H.; Li, X.; Huang, W.; Li, C. Chair Style Recognition and Intelligent Design Method Based on Deep Learning. *Furniture* **2021**, *42*, 37–41.
10. Qin, B.; Gu, N.; Zhang, X. Image verification code recognition based on convolutional neural network. *Comput. Syst. Appl.* **2018**, *27*, 144–150.
11. Yan, X.; Huang, S. Lightweight target detection network based on grouped heterogeneous convolution. *Comput. Sci.* **2020**, *47*, 114–117.
12. Cao, J.; Li, Y.; Sun, M.; Yi, C.; Lischinski, D.; Cohen-Or, D.; Chen, B.; Tu, C. DO-Conv: Depthwise Over-parameterized Convolutional Layer. *IEEE Trans. Image Process.* **2022**, *31*, 3726–3736. [[CrossRef](#)] [[PubMed](#)]
13. Xie, S.; Zheng, X.; Chen, Y.; Xie, L.; Liu, J.; Zhang, Y.; Yan, J.; Zhu, H.; Hu, Y. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction. *Sci. Rep.* **2018**, *8*, 1–9. [[CrossRef](#)] [[PubMed](#)]