AUTOMATED AND LIGHTWEIGHT NETWORK DESIGN VIA RANDOM SEARCH FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

Jihao Li [1, 2, 3], Wenhui Diao [1, 2, *], Xian Sun [1, 2, 3], Yingchao Feng [1, 2, 3], Wenkai Zhang [1, 2], Zhonghan Chang [1, 2, 3], Kun Fu [1, 2, 3]

[1] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[2] Key Laboratory of Network Information System Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China - (whdiao, sunxian, fukun)@mail.ie.ac.cn, zhangwk@aircas.ac.cn
[3] University of Chinese Academy of Sciences, Beijing, China - (lijihao17, fengyingchao17, changzhonghan16)@mails.ucas.edu.cn

KEY WORDS: Scene Classification, Random Search, Neural Architecture Search, Remote Sensing Image, Deep Learning, Semantic Segmentation

ABSTRACT:

With the development of deep learning, remote sensing image scene classification technology has been greatly improved. However, current deep networks used for scene classification usually introduce ingenious extra modules to fit the characteristics of remote sensing images. It causes a high labor cost and brings more parameters, which makes the network more complicated and poses new intractable problems. In this paper, we rethink this popular "add module" pattern and propose a more lightweight model, called ProbDenseNet (PDN). PDN is obtained via a random search strategy in Neural Architecture Search (NAS) which is an automated network design manner. In our method, all topological connections are assigned importance degrees which subject to a uniform distribution. And we set a regulator to adjust the sparsity of the network. By this way, the design procedure is more automated and the network structure becomes more lightweight. Experimental results on AID benchmark demonstrate that the proposed PDN model can achieve competitive performance even with much fewer parameters. And we also find that excessive connections do not always improve the network's performance while they can drag down the network's behavior as well. Furthermore, we conduct experiments on Vaihingen dataset with classical Fully Convolutional Network (FCN) framework. Quantitative and qualitative results both indicate that the features learned by PDN can also transfer in semantic segmentation task.

1. INTRODUCTION

Remote sensing image scene classification is a fundamental work not only in Computer Vision but also in Earth Vision (Cheng, Han, 2016, Xia et al., 2018). The purpose of remote sensing image scene classification is to efficiently and automatically identify the semantic category label through some algorithms. It has a significant impact on Land Use and Land Cover (LULC) determination (Zhang et al., 2013, Zhu et al., 2016), vegetation mapping (Li, Shao, 2013, Mishra, Crews, 2014), urban planning and so on. While it also offers a foundation for semantic segmentation (Kampffmeyer et al., 2016), object detection (Wang et al., 2019a, Fu et al., 2020, Feng et al., 2019), Fine Grained Visual Classification (FGVC) (Fu et al., 2019) and other extension tasks. Due to the rapidly increasing quantity of remote sensing images, highly complex geometric structures and quite large scale images (Zhao et al., 2016), how to improve a model's performance and increase its automation of design procedure are still tricky problems.

Before the booming of deep learning, low-level feature based methods (Penatti et al., 2015) and middle-level feature based methods (Zhong et al., 2015, Zhao et al., 2015) are the major avenues to deal with the remote sensing image scene classification task. Since 2012, deep learning methods or high-level feature based methods represented by AlexNet (Krizhevsky et al., 2012) have shown extraordinary talents in lots of visual tasks including remote sensing image scene classification (Zhu et al., 2017). Among numerous deep learning methods, Convolutional Neural Networks (CNNs) are widely used. They have a strong interpretability in neurological theory and these architectures are pretty adept at various image processing tasks. Therefore, the amazing results immediately roused a huge upsurge of study in them.

(a) Orientation is quite different.

(b) Scale changes tremendously.

Figure 1. Some challenges in remote sensing scenes. (a) The direction of a baseball field can be downward or upward. (b) A bridge can be as small as several thousand pixels and even as large as millions of pixels.

* Corresponding author

However, remote sensing images are extremely different from natural scene images. Under the influence of solar elevation angles, flying altitude, etc., the appearance of remote sensing image scenes may vary significantly. Figure 1 illustrates some challenges in remote sensing images. Therefore, it brings great difficulties in designing a network which aims at solving the problem of remote sensing image scene classification. Current deep learning methods usually leverage some ingenious extra structures to enhance the original CNNs' capability of feature expression and feature extraction, such as MIDCCNN (Bi et al., 2019) and ResNet-TP (Zhou et al., 2018). This is more like an "add module" pattern. The design process of these ingenious modules usually needs strong expert knowledge and expert experience in remote sensing. Even so, a series of trial and error is also inevitable and it is still an extremely laborious task. Meanwhile, the quantity of parameters also increases necessarily. Complex network structure and high GPU memory usage will pose new intractable problems as well.

Recently, a novel design paradigm of neural networks known as Neural Architecture Search (NAS) (Zoph, Le, 2016, Zoph et al., 2018) or AutoML (Quanming et al., 2018, Xie et al., 2019) theory has attracted much attention. This method can simplify the process of designing new network architectures. In remote sensing field, it also has a wide utilization (Chen et al., 2019, Bui et al., 2018). In this paper, enlightened by this automated manner, we rethink the effect of the popular "add module" pattern and focus on an automated and lightweight network design procedure in remote sensing image scene classification task. We deem that this "add module" pattern, as mentioned above, is just a remedy to neutralize the negative effects of redundant connections in some extent. Excessive connections can degrade the networks' performance, just like opposite resultant force component. Utilizing reasonable measures to properly eliminate useless branches probably obtains a very good effect as well. Consequently, we propose a simple but effective random strategy to search a compact neural network model, ProbDenseNet (PDN) based on DenseNet (Huang et al., 2017). This approach is to live up to the full potential of key parts. It is much more like a "sub module" pattern. The topological connection relations can be cut off automatically to search key parts from the perspective of machine itself. It can lower the threshold for network architecture design. Our main works can be briefly summarized as follows:

1. A novel "sub module" pattern is proposed through a simple and effective random search strategy to prune excessive connections automatically. And this evaluation process which determines the importance of connection relationships is all standing on the perspective of machine itself.
2. Considering the flexibility of this method, we set a regulator to adjust the sparsity of the network conveniently. It can be more flexible to fit various application environments under the control of regulator.
3. The proposed PDN model achieves the highest classification metrics with minimum parameters both on 20% and 50% AID dataset (Xia et al., 2017). And we also find that excessive connections do not always improve the network's performance while they can drag down the network's behavior as well.
4. We further test PDN's performance in semantic segmentation task with classical Fully Convolutional Network (FCN) (Long et al., 2015) framework on Vaihingen dataset[1]. Comparative results indicate that the features learned by PDN model can also transfer in different visual tasks.

[1] http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html

2. RELATED WORK

2.1 CNNs for Remote Sensing Image Scene Classification

Since AlexNet (Krizhevsky et al., 2012) won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Deng et al., 2009) championship in a landslide, Convolutional Neural Networks (CNNs) have become the most popular method in a variety of visual tasks, such as scene classification, semantic segmentation and object detection. In order to enhance the original CNNs' feature extraction capability, some ingenious extra modules are designed to attach to existing popular networks when processing remote sensing images. (Zhou et al., 2018) propose an elaborate two-pathways module on the basis of ResNet (He et al., 2016) to aggregate both local details and regional context of remote sensing images. Different from most structures in scene classification, dilation convolution (Yu, Koltun, 2015) is greatly used in this model. MIDCCNN (Bi et al., 2019) adds an attention-based multiple instance pooling structure on original DenseNet (Huang et al., 2017) to highlight the local semantics in remote sensing scenes. Furthermore, (Wang et al., 2019b) introduce a sibling network for feature embedding of remote sensing images to boost the classification performance. These ingenious extra modules bring a large number of parameters and make networks more complicated. These methods do not realize the full potential of original CNNs. Moreover, the reliance on strong expert knowledge in remote sensing and this laborious design procedure also limit the wide application of these methods.

2.2 Neural Architecture Search

Current Neural Architecture Search (NAS) framework commonly employs Reinforcement Learning (RL) algorithm to search network architectures. In these methods, RL is regarded as an optimization strategy. Neural networks are defined as a digital sequence which is generated by a controller, generally a Long Short Term Memory (LSTM) (Hochreiter, Schmidhuber, 1997). Next, the network derived by this sequence is trained in order to return an accuracy on validation set. Then, this accuracy serves as a reward to update the controller through Policy Gradient (PG) (Sutton et al., 2000). When the search process is complete, the derived model's weights learned in search stage will be discarded and they will be trained from scratch. As regarding to the automated design paradigm of CNNs for remote sensing images, there are also some applications. For instance, (Chen et al., 2019) firstly apply NAS framework to Hyperspectral Image (HSI) classification and propose 1-D Auto-CNN and 3-D Auto-CNN as HSI classifiers. Both of these models obtain state-of-the-art performance on four public hyperspectral datasets. Although the success of NAS inspires lots of valuable works, extremely expensive computational cost (28 days with 800 GPUs) (Zoph et al., 2018) still makes most researchers daunted.

In this paper, we do not leverage the popular "add module" pattern to design a neural network; on the contrary, we discard some topological structures since the redundancy and we fully exploit the potential of the network. By means of NAS framework, the design procedure can become more automated. Topological connection relations are determined by machine itself, rather than manual work. And our random search strategy is

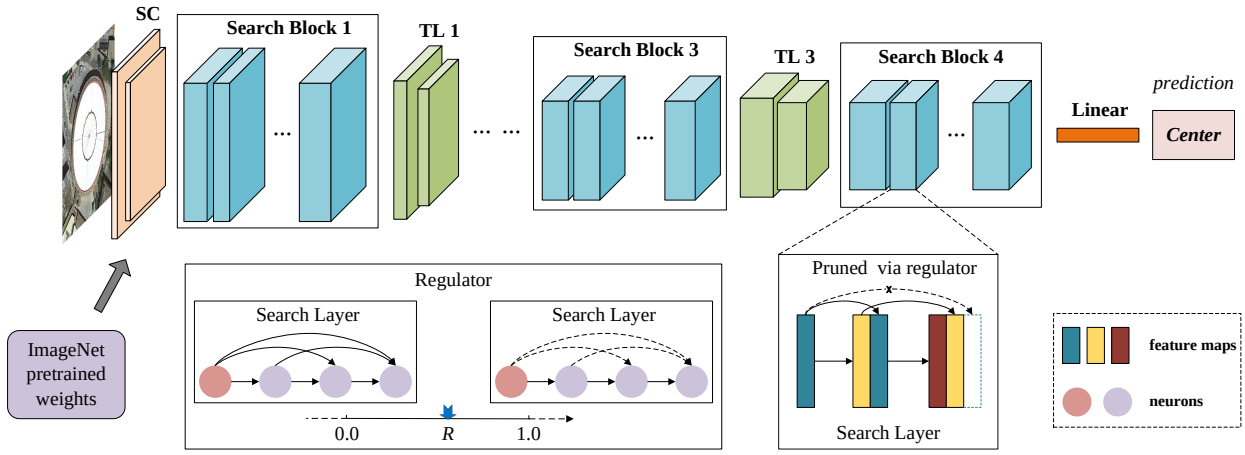


Figure 2. The whole pipeline of the proposed PDN model and the diagram of regulator. SC means Stem Convolution and TL means Transition Layer. Here, we omit some architectures of PDN just for convenience. In PDN model, the sparsity of Search Blocks is adjusted through a regulator. The value of regulator *R* is closer to 0.0, Search Blocks are more complex. On the contrary, Search Blocks are sparser when *R* is closer to 1.0. All Search Blocks are controlled by this regulator while Stem Convolution and Transition Layers stay the same. Besides, Linear Layer is consistent with the number of categories. Best view in color.

also simple and effective. In our method, the design threshold is reduced greatly while the generated model becomes more lightweight as well.

3. METHODOLOGY

In this section, we introduce a simple and automated design procedure to prune excessive branches. Not only are parameters reduced, but also the dependence on expert knowledge in remote sensing is decreased. First, the structure of original DenseNet is reviewed briefly in section 3.1. Then the pruning process is explained detailedly in section 3.2.

3.1 A Review of DenseNet

Original DenseNet (Huang et al., 2017) consists of 1 Stem Convolution, 4 Dense Blocks, 3 Transition Layers and 1 Classification Layer. Numerous parameters are attached to Dense Blocks which consist of a series of 1×1 convolution and 3×3 convolution. The forward propagation process in a Dense Block can be computed through Equation 1:

$$x_n = \mathbb{F}_n([x_0, x_1, \cdots, x_{n-1}]) \tag{1}$$

where, x_n is the n-th layer's feature map. $\mathbb{F}_n$ indicates a mapping of this layer. $[x_0, x_1, \cdots, x_{n-1}]$ denotes a concatenation operation which can merge previous layers' feature map. Although this combination may be beneficial to feature fusion, it also brings the growth of parameters. And we find that this growth may reduce the performance of the network through experiments. Hence, we need to cut off some unimportant connections to compress the network. The process of how to prune the network will be explained in the next subsection.

3.2 The Pruning Process

The macro-architecture of PDN model is described in Figure 2. In PDN model, Dense Blocks are pruned through search method. So we call it Search Blocks. Other configurations of PDN are similar to original DenseNet121. A PDN model takes as input a remote sensing image. PDN first processes the image with a Stem Convolution. Next, Search Blocks extract semantic information of this image and Transition Layers adjust the size of feature maps in turn. Then, a prediction label is produced through a Linear Layer. Finally, we use Cross Entropy (CE) loss function to minimize the loss between predictions and the ground truth labels.

In PDN model, the forward propagation of a Search Block can be written by:

$$x_n = \mathbb{F}_n([\mathbb{1}(x_0), \mathbb{1}(x_1), \cdots, \mathbb{1}(x_{n-1})]) \tag{2}$$

where $\mathbb{1}(\bullet)$ is an indicator function which indicates whether the feature map of a previous layer is merged or not. Different from Dense Blocks in original DenseNet, we carefully select which feature maps are concatenated, rather than merge all feature maps aimlessly. Therefore, the network complexity can be reduced after properly discarding some unimportant feature maps. Note that other types of layers remain unchanged, i.e. Stem Convolution and Transition Layers, except Linear Layer which will be changed according to the number of categories in a dataset.

In order to reduce manual work and the dependence on expert knowledge in remote sensing, we utilize a random search approach to prune automatically. We assign each layer's feature map an importance degree which can distinguish the role of different layers in a Search Block. With the flow of data, the closer a layer is to the output in a Search Block, the more importance degrees a layer owns. Therefore, these importance degrees form an upper triangular matrix. For the sake of fairness and randomness, we let these importance degrees subject to a uniform distribution. It can be expressed as:

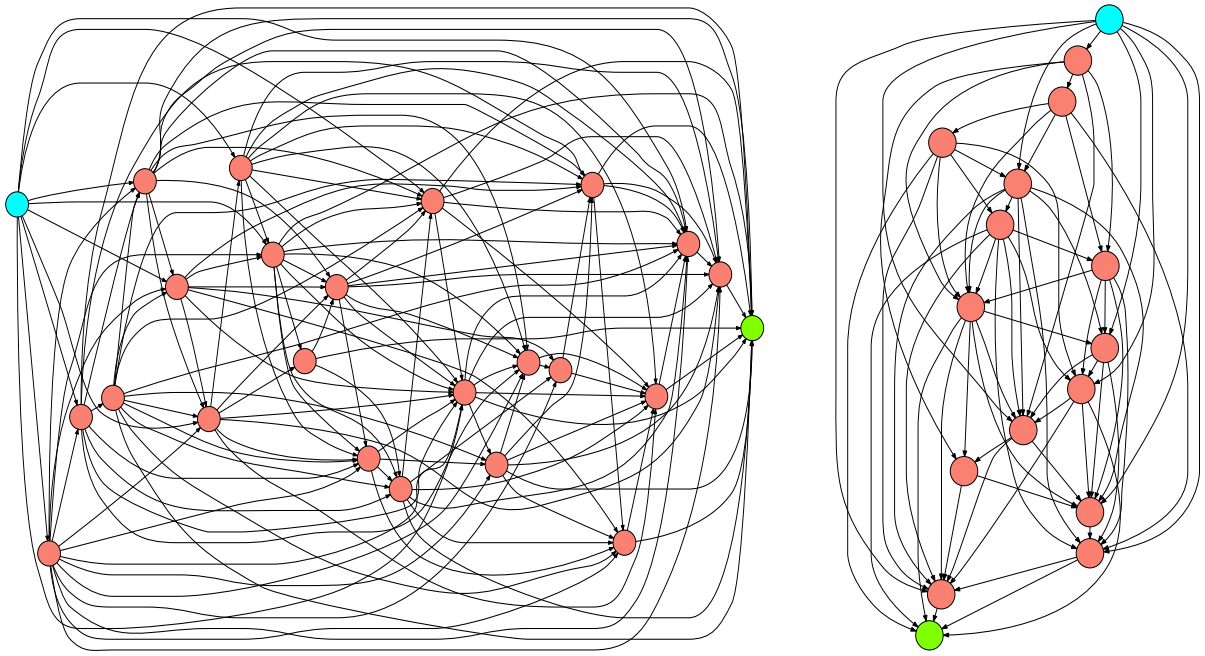
$$\begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \cdots & \theta_{0,n-1} \\ 0 & \theta_{1,1} & \cdots & \theta_{1,n-1} \\ \vdots & \ddots & \vdots & \theta_{2,n-1} \\ 0 & \cdots & 0 & \theta_{n-1,n-1} \end{bmatrix} \sim uniform(0,1) \tag{3}$$

where θ represents the importance degree attached to a certain layer in a Search Block.

Then we set a regulator to determine which connections are cut off. The diagram for adjusting the sparsity of PDN is illustrated

Network	Method	OA(%), 20%	OA(%), 50%	Params(MB)	Speed(ms)
GIST (Xia et al., 2017)	Low	30.61±0.63	35.07±0.41	-	-
BoVW (CH) (Xia et al., 2017)	Mid	48.60±0.41	55.74±0.48	-	-
VLAD (LBP) (Xia et al., 2017)	Mid	59.44±0.43	69.42±0.85	-	-
IFK (SIFT) (Xia et al., 2017)	Mid	70.60±0.42	77.33±0.37	-	-
AlexNet (Xia et al., 2017)	High	86.86±0.47	89.53±0.31	57.13	**3.70**
VGGNet (Xia et al., 2017)	High	86.59±0.29	89.64±0.36	134.38	38.47
GoogLeNet (Xia et al., 2017)	High	83.44±0.40	86.39±0.55	10.02	18.91
DCCNN (Bi et al., 2019)	High	-	91.49±0.22	5.35	65.43
MIDCCNN (Bi et al., 2019)	High	-	92.53±0.18	7.48	117.22
PDN (0.3)	NAS	**89.82±0.25**	**94.43±0.26**	**4.90**	4.95

Table 1. Comparison results of the proposed ProbDenseNet (0.3) on 20% and 50% AID respectively.



(a) Search Block 3 (b) Search Block 4

Figure 3. Visualization results of Search Block 3 and Search Block 4 in PDN moel. Cyan, salmon and chartreuse represent input node, middle node and output node respectively. The closer a node is to the output node, the more connections a node may own. Topological connections are all determined by machine itself.

in Figure 2. When the importance degree is greater than the value of regulator R, the branch it attaches to will be kept. On the contrary, the branch will be cut off. The function of this regulator is more like a High Pass Filter (HPF). So the smaller R is, the heavier the network will be, and vice versa. Under the effect of R, the network pruning process can become more flexible to fit various application environments. When R is decided, the network architecture is also fixed.

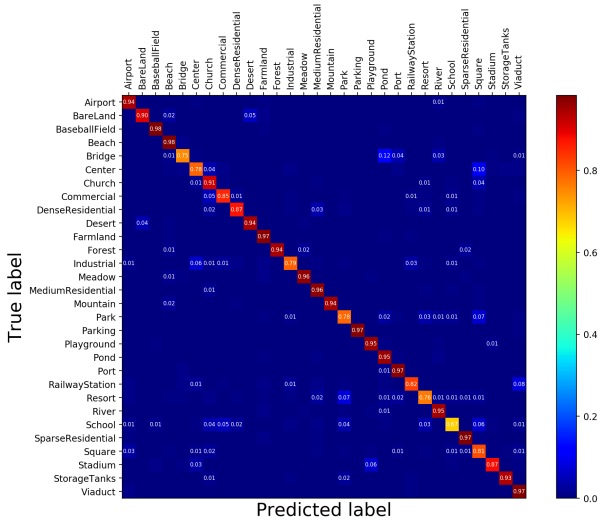
4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we describe the steps of experiments and analyze corresponding results to confirm the effectiveness of PDN model. First, we separately perform experiments on 20% and 50% AID dataset (Xia et al., 2017) to verify its capability of feature extraction in section 4.1. And, in section 4.2, we conduct experiments in semantic segmentation task with classical Fully Convolutional Network (FCN) (Long et al., 2015) framework on Vaihingen dataset to show that the features learned by PDN can also transfer.

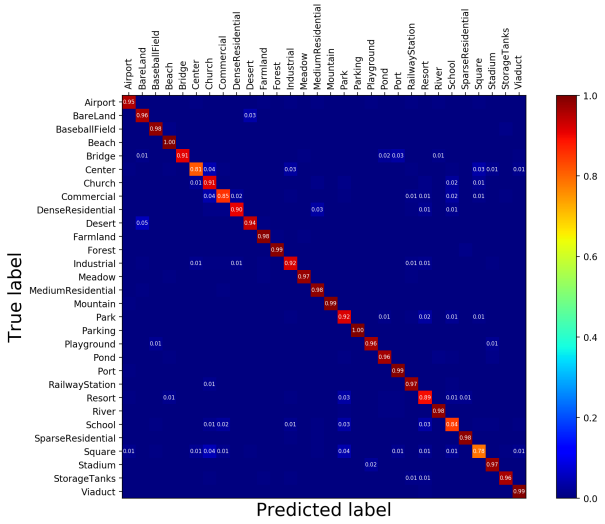
4.1 Results on Scene Classification

4.1.1 AID Dataset We confirm the validity of the proposed PDN model on AID which is widely used in remote sensing image scene classification task. Higher intraclass variations, smaller interclass dissimilarity and relative large scale make AID very challenging. There are 10,000 remote sensing images dispersed in 30 categories. The number of images in each category varies from 220 to 420. Samples of each category are unbalanced. The size of each image is 600×600 and the spatial resolution changes from 8m to 0.5m. According to the official statement on this benchmark, the ratio of training set is fixed to be 20% and 50%, and the left as testing samples.

4.1.2 Implementation Details To accelerate convergence, we utilize transfer learning strategy which initializes PDN model by loading the pretrained weights on ILSVRC (Deng et al., 2009). For optimization strategy, we adopt momentum Stochastic Gradient Descent (SGD) (Qian, 1999) with a cosine schedule where the learning rate anneals down from 0.05 to



(a) Overall accuracy (OA) on 20% AID



(b) Overall accuracy (OA) on 50% AID

Figure 4. Confusion matrices of the best results achieved by PDN model on 20% and 50% AID respectively. Note that red represents better classification performance while blue means worse. Best view in color.

10^{-6}. Meanwhile, considering the over-fitting problem, weight decay and dropout rate is set to be 5×10^{-3} and 0.2. Besides, the number of iterations is 120 and 150 on 50% AID and 20% AID respectively. To realize an objective and comprehensive evaluation, we compare various methods by computing Overall Accuracy (OA), the number of parameters and the inference speed respectively which are widely used in image classification task.

4.1.3 Comparison with Baselines The results of comparison experiments are listed in Table 1. Here, we compare some famous low-level, middle-level and high-level feature based methods. And the value of regulator R is set to be 0.3. Figure 3 shows two topology diagrams in PDN model: Search Block 3 and Search Block 4.

From these comparison results, we can see that our proposed PDN model obtained by automatically random "sub module" pattern method achieves the highest OA with minimum parameters both on 20% and 50% AID dataset. It is superior to all baseline models, especially for high-level feature based methods, such as VGGNet (Simonyan, Zisserman, 2014) and GoogLeNet (Szegedy et al., 2015). And the performance of inference speed is also not inferior. In addition, it is better than the "add module" pattern model MIDCCNN (Bi et al., 2019) in all aspects. While the convergence speed is accelerated as well; for instance, 300 epochs are required in training DCCNN (Bi et al., 2019) and MIDCCNN (Bi et al., 2019), but merely 120 epochs are needed in training PDN model on 50% AID. More importantly, it should be emphasized that the design process of PDN model does not involve plenty of expert knowledge and artificial factors. It is just an automated procedure and makes full use of the potential of key parts.

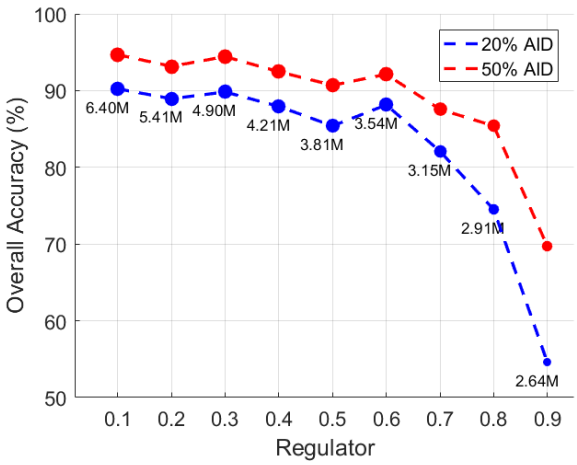


Figure 5. The classification performance of different regulator settings on 20% and 50% AID dataset. The number of parameters is marked below the blue curve.

Besides giving these performance metrics as mentioned previously, we also compute corresponding confusion matrices on the best results of PDN model, which are shown in Figure 4. These two confusion matrices are obtained on 20% and 50% AID separately. It can be seen that remote sensing scene types can be distinguished easily on 50% AID. And the accuracies of most categories are up to 0.9. Particularly, two scenes, namely beach and parking, reach 1.0. These two kinds of scenes are all correctly classified. As for the confusion matrix on 20% AID, the performance of PDN is not as outstanding as that on 50% AID due to lack of training samples. The accuracy of 6 scenes, i.e. bridge, center, industrial, park, resort and school, is less than 0.8. However, for most scenes, the classification results are yet acceptable. Most values of this confusion matrix are concentrated on the principal diagonal. And the accuracy of the rest 24 categories are all higher than 0.9. In general, it is also a very competitive behavior in the case of quite few training samples.

4.1.4 The Analysis of Different Regulator Settings Figure 5 illustrates the impact of different regulator settings on classification performance. We set the value of regulator in an interval from 0.1 to 0.9. Then, we record the accuracy and the number of parameters every a fixed interval of 0.1. It can be observed that accuracy is not proportional to parameters both on 20% and 50% AID. More parameters do not mean better performance, even having an opposite effect. For instance, parameters at $R = 0.2$ is more than that at $R = 0.3$, but the accuracy at $R = 0.2$ is worse than $R = 0.3$. Meanwhile, we also find that

Network	Imp. surf.	Build.	Low veg.	Tree	Car	mean F_1	mIoU	OA
FCN32 (Long et al., 2015)	87.54	91.83	75.56	84.74	49.45	77.83	65.97	85.02
FCN32-PDN (0.3)	88.83	**94.02**	77.55	86.41	64.86	82.33	71.20	86.88
FCN16 (Long et al., 2015)	87.44	91.53	77.28	86.40	70.16	82.56	71.02	85.88
FCN16-PDN (0.3)	**89.78**	93.86	**78.52**	85.44	76.42	84.80	74.19	86.91
FCN8 (Long et al., 2015)	85.75	89.80	75.86	86.71	78.52	83.33	71.77	84.91
FCN8-PDN (0.3)	88.58	92.69	78.44	**88.07**	**81.25**	**85.81**	**75.50**	**87.28**

Table 2. Quantitative results on Vaihingen by using FCN framework with different upsampling strides.

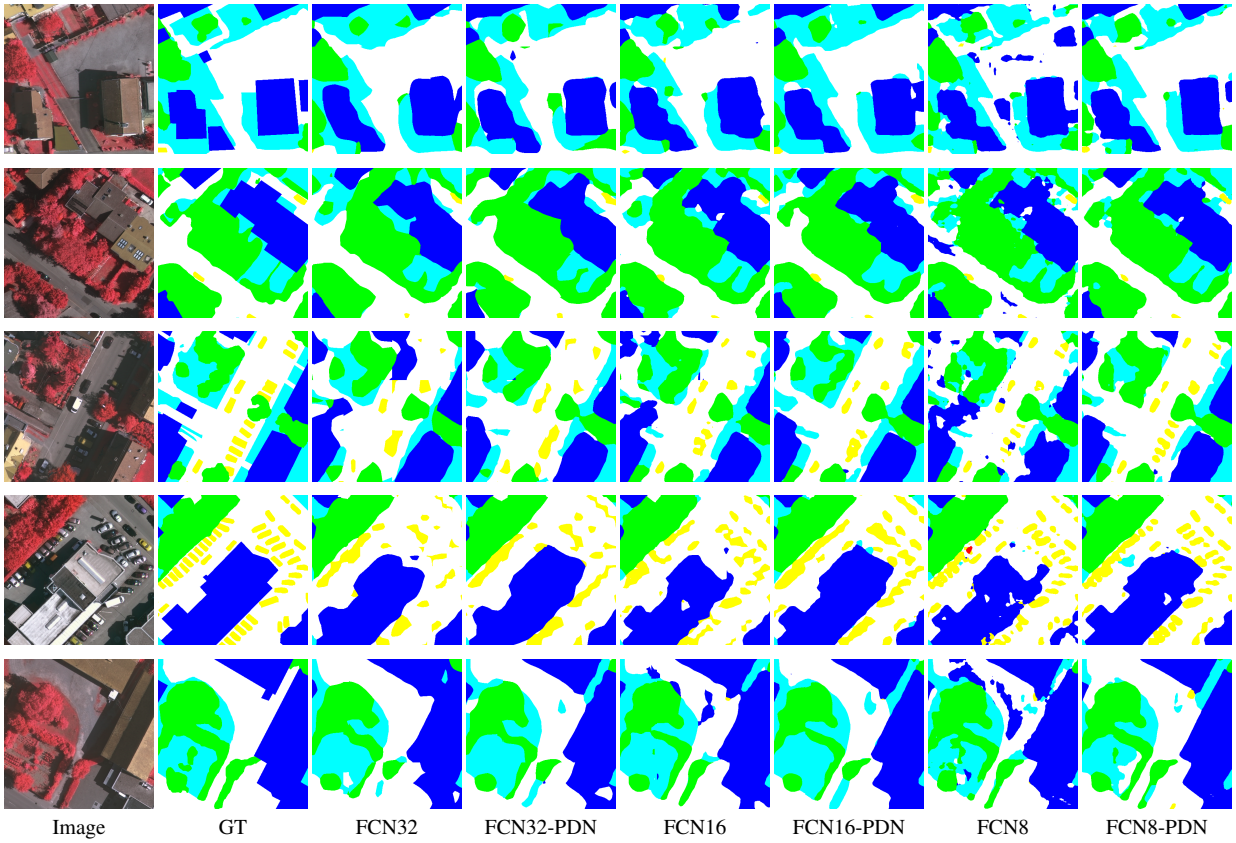


Figure 6. Visualization results on Vaihingen dataset. The first and second columns separately present raw images and their corresponding ground truth labels. The remaining columns show the comparison results with the original FCNs according to the upsampling strides of 32, 16 and 8. Best view in color.

the accuracy metric varies little when $R \leq 0.6$. The performance of each other is very close. But if $R > 0.6$, the accuracy drops sharply. And it is notable that $R = 0.3$ for the proposed PDN model is the best trade off between accuracy and the number of parameters.

4.2 Vaihingen Semantic Segmentation

4.2.1 Vaihingen Dataset Vaihingen dataset contains 33 large scale images with an average size of 2494×2064 pixels. These images are collected over a 1.38 km^2 area of Vaihingen, a city in Germany. The spatial resolution is 9 cm. And each image has three different bands: corresponding to near infrared, red and green wavelengths. Among all these 33 images, only 16 of them are provided with pixel-wise ground truth labels which include impervious surfaces, buildings, low vegetation, trees, cars and clutter/background. Following the previous works (Volpi, Tuia, 2016, Maggiori et al., 2017, Marcos et al., 2018), 11 images are selected for training and the rest 5 images (namely 11, 15, 28, 30, 34) are used for testing.

4.2.2 Implementation Details As (Zoph et al., 2018) did to show the transfer capability of the proposed model, we plug the PDN (0.3) into the classical FCN framework to conduct semantic segmentation experiments and compare the original FCN in accordance with their upsampling strides, i.e. 32, 16, 8. Regarding the experimental settings, we adopt SGD optimizer with a batch size of 6. And to smooth the gradients, we set the value of momentum to be 0.9. While considering the over-fitting problem, weight decay is also utilized and set to be 10^{-3}. We crop the size of input samples to 513×513 and train all models for 200 epochs. During the training stage, the learning rate which is initialized to 0.01 decays as a poly scheduler with the power of 0.9. The evaluation follows three widely accepted protocol of mean F_1 score, mean Inter-section over Union (mIoU) and Overall Accuracy (OA).

4.2.3 Comparison Results Table 2 shows quantitative results on Vaihingen dataset. We can observe that FCNs with PDN (0.3) as the backbone surpass all three original FCNs in terms of these three metrics we adopt. Specifically, there is a significant

improvement on small size objects, such as cars. The F_1 score of car category tested on the original FCN32 is merely 49.45. But on FCN32-PDN (0.3), this metric can achieve 64.86, about a 31.16% increments. And the performance of FCNs which are combined with PDN (0.3) is also improved, when upsampling strides are 16 and 8. During experiments, we found that it is easy to be over-fitting for the original FCNs. Nevertheless, the performance of FCN-PDNs on test set is quite equal to that on training set. This also indicates that the proposed PDN is a very robust model. Figure 6 presents some visualization results on Vaihingen dataset. It can be seen that the boundary information and detail information are more obvious in FCN-PDNs. Hence, we can conclude that the proposed PDN model can provide superior and generic remote sensing imagery features. And these features extracted in remote sensing image scene classification can also transfer in semantic segmentation task. That is a powerful demonstration of feature transfer capability of the proposed PDN model.

5. CONCLUSION

In this paper, we have rethought the popular network design process of "add module" pattern in remote sensing image scene classification task. This pattern does not exploit the full potential of the network in some extent. An extra module may not improve the network's performance. Instead, this brings more parameters and makes the network more complicated, which poses new intractable problems in the training stage. Meanwhile, the design procedure is laborious and heavily relies on strong expert knowledge. Therefore, we propose a "sub module" pattern to design a neural network, ProbDenseNet (PDN) on the basis of recently presented NAS manner. In our method, the generation process can be executed according to the assigned importance degrees via a simple but effective random search strategy. It is an automated procedure, rather than manual work. The threshold for network architecture design is lowered as much as possible and the generated model can become more lightweight. Experiments on AID dataset show that the PDN model can get better classification performance even with quite few parameters. It strongly verifies the feature extraction capability of PDN. And we found that the number of parameters is not proportional to model's performance. It suggests that excessive connections do not always improve the network's performance while they can drag down the network's behavior as well. Besides, the results on Vaihingen dataset demonstrate PDN's high feature transfer capability in semantic segmentation task. FCNs with PDN (0.3) as the backbone surpass all baseline models in terms of mean F_1 score, mIoU, and OA metrics. As for future works, we will try to apply the PDN model on larger and more complex datasets, such as multispectral remote sensing images, and more visual tasks to evaluate it deeply.

6. ACKNOWLEDGEMENTS

The authors would greatly appreciate all anonymous reviewers for their valuable comments which are of benefit for us. This work is supported by the National Natural Science Foundation of China (No. 41701508).

REFERENCES

Bi, Q., Qin, K., Li, Z., Zhang, H., Xu, K., 2019. Multiple instance dense connected convolution neural network for aerial image scene classification. *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2501–2505.

Bui, D., Tran, T., Nguyen, T., Tran, Q., Van Nguyen, D., 2018. Aerial Image Semantic Segmentation Using Neural Search Network Architecture. 113–124.

Chen, Y., Zhu, K., Zhu, L., He, X., Ghamisi, P., Benediktsson, J. A., 2019. Automatic Design of Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 7048–7066.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 11–28.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 248–255.

Feng, Y., Diao, W., Sun, X., Yan, M., Gao, X., 2019. Towards Automated Ship Detection and Category Recognition from High-Resolution Aerial Images. *Remote Sensing*, 11(16), 1901.

Fu, K., Chang, Z., Zhang, Y., Xu, G., Zhang, K., Sun, X., 2020. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161, 294–308.

Fu, K., Dai, W., Zhang, Y., Wang, Z., Yan, M., Sun, X., 2019. Multicam: Multiple class activation mapping for aircraft recognition in remote sensing images. *Remote Sensing*, 11(5), 544.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, 9(8), 1735–1780.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708.

Kampffmeyer, M., Salberg, A.-B., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1–9.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097–1105.

Li, X., Shao, G., 2013. Object-based urban vegetation mapping with high-resolution aerial photography as a single data source. *International journal of remote sensing*, 34(3), 771–789.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 7092–7103.

Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS journal of photogrammetry and remote sensing*, 145, 96–107.

Mishra, N. B., Crews, K. A., 2014. Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with Random Forest. *International Journal of Remote Sensing*, 35(3), 1175–1198.

Penatti, O. A., Nogueira, K., Dos Santos, J. A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 44–51.

Qian, N., 1999. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1), 145–151.

Quanming, Y., Mengshuo, W., Hugo, J. E., Isabelle, G., Yi-Qi, H., Yu-Feng, L., Wei-Wei, T., Qiang, Y., Yang, Y., 2018. Taking human out of learning applications: A survey on automated machine learning. *arXiv preprint arXiv:1810.13306*.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sutton, R. S., McAllester, D. A., Singh, S. P., Mansour, Y., 2000. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 1057–1063.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Volpi, M., Tuia, D., 2016. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 881–893.

Wang, P., Sun, X., Diao, W., Fu, K., 2019a. FMSSD: Feature-Merged Single-Shot Detection for Multiscale Objects in Large-Scale Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*.

Wang, W., Du, L., Gao, Y., Su, Y., Wang, F., Cheng, J., 2019b. A discriminatively learned cnn embedding for remote sensing image scene classification. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3832–3835.

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. Dota: A large-scale dataset for object detection in aerial images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3974–3983.

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 3965–3981.

Xie, S., Kirillov, A., Girshick, R., He, K., 2019. Exploring randomly wired neural networks for image recognition. *arXiv preprint arXiv:1904.01569*.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhang, Y., Sun, X., Wang, H., Fu, K., 2013. High-resolution remote-sensing image classification via an approximate earth mover's distance-based bag-of-features model. *IEEE Geoscience and Remote Sensing Letters*, 10(5), 1055–1059.

Zhao, B., Zhong, Y., Xia, G.-S., Zhang, L., 2015. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4), 2108–2123.

Zhao, B., Zhong, Y., Zhang, L., 2016. A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 73–85.

Zhong, Y., Zhu, Q., Zhang, L., 2015. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11), 6207–6222.

Zhou, Z., Zheng, Y., Ye, H., Pu, J., Sun, G., 2018. Satellite image scene classification via convnet with context aggregation. *Pacific Rim Conference on Multimedia*, Springer, 329–339.

Zhu, Q., Zhong, Y., Zhao, B., Xia, G.-S., Zhang, L., 2016. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 13(6), 747–751.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.

Zoph, B., Le, Q. V., 2016. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.

Zoph, B., Vasudevan, V., Shlens, J., Le, Q. V., 2018. Learning transferable architectures for scalable image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8697–8710.