Deep Convolutional Neural Network Based Image Segmentation for Salt Mine Recognition

Ming Tao[1(✉)] 🆔, Xueqiang Li[1,2] 🆔, and Kai Ding[1] 🆔

[1] School of Computer Science and Technology,
Dongguan University of Technology, Dongguan 523808, People's Republic of China
ming.tao@mail.scut.edu.cn, {lixq,dingkai}@dgut.edu.cn
[2] Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou 510006,
People's Republic of China

Abstract. Scientifically, the identification of salt ore has definite practical significance for the exploitation of oil and gas. Traditionally, this is achieved by picking the salt boundaries with manual vision, which may introduce serious systematic bias. Nowadays, with the technological progress of machine vision used in image analysis, human effort has been replaced by machine capacity in salt mine recognition. Especially, with the in-depth application of deep learning technology in the field of machine vision, salt mine recognition using image analysis is revolutionizing with more acceptable efficiency and accuracy. To this end, with exploratory data analysis to mine the characteristics and data processing to increase the size of the image data for further enhancing the generalization capability of the designed model, a deep convolutional neural network based image segmentation model is investigated to achieve salt mine recognition in this paper. Concretely, a U-Net model integrated modified ResNet34 is first designed as a basic recognition model, and many attempts then are conducted to further optimizing the model according to the data characteristics, including adding auxiliary function, hyper-column, scSE and depth supervision scheme. In addition, multiple loss functions are also attempted to be adapted to further improving the model generalization capacity. The numerical analysis and evaluation finally show the efficiency of the investigations on loss value and recognition accuracy.

Keywords: Deep learning · Convolutional neural network · Image segmentation · Salt mine recognition

1 Introduction

Science and practice show that large deposits of oil and gas also tend to form large deposits of salt beneath the earth's surface. Therefore, we can use the search for salt mines to explore the locations of oil and gas. In general, determining and segmenting the locations of large salt mines are to regionally segment the instance of salt mine and take the rest as foreground, which can be achieved by analyzing seismic images through very professional manual vision, which are subjective and inefficient. Fortunately, with the technological progress of machine vision used in image analysis, machine capacity

has gradually replacing human effort in salt mine recognition, which could provide higher efficiency and accuracy. Nowadays, due to offering a convincing alternative by learning the problem specific features on their own, the in-depth development of deep learning technology in the field of image analysis has further promoting the application of machine vision in various applications, e.g., medical image segmentation, salt mine recognition, etc. In particular, as a frequently used deep learning framework, convolutional neural networks (CNN) is usually used for achieving image segmentation [1].

For accurate medical image segmentation used in diagnosis, surgical planning and many other applications, Wang et al. [2] proposed a deep learning-based interactive segmentation method to improve the results obtained by an automatic CNN and to reduce user interactions during refinement for higher accuracy. With the facilitated gradient flow and implicit deep supervision during training achieved by DenseNet, Dolz et al. [3] proposed HyperDenseNet as a 3-D fully CNN that extended the definition of dense connectivity to multi-modal segmentation problems. In [4], the great potential of applying CNN for salt-related interpretations was also demonstrated. Concretely, the benefits of CNN-based classification are demonstrated by using U-Net network structure, along with the residual learning framework ResNet, to delineate salt body with high precision.

To this end, a method of deep convolutional neural network based image segmentation is proposed for achieve salt mine recognition in this paper. Concretely, a U-Net model integrated modified ResNet34 is first designed as a basic recognition model, and many attempts then are conducted to further optimizing the model according to the data characteristics, and multiple loss functions are also attempted to be adapted to further improving the model generalization capacity.

The rest of this paper is organized as follows. In Sect. 2, exploratory data analysis and processing are conducted to mine the data characteristics and to increase the size of the image data for further enhancing the generalization capability of the designed model. In Sect. 3, a U-Net model integrated modified ResNet34 is designed and many attempts are conducted to further optimizing the model according to the data characteristics. In Sect. 4, multiple loss functions are attempted to be adapted in the salt mine image segmentation model to further improve the model generalization capability. In Sect. 5, numerical analyses are conducted to evaluate the efficiency of achieved proposals in this work. In Sect. 6, this paper is summarized and concluded.

2 Exploratory Data Analysis and Processing

The image data source for salt mine recognition is seismic images collected by the seismic reflection. In this work, the image data source is provided by GS Salt Identification Challenge conducted in Kaggle, where, the train.zip is the training set containing 4000 sample images and the corresponding masks, the test.zip contains 18000 testing sample images, the train.csv contains the sample images' IDs and the labels of salt in the training set, and the depth.csv contains the depth of the sample images. To better understand the data and mine the characteristics of the data, which may play a certain role in the optimization of the designed salt mine recognition model, salt mine coverage rate of samples in the training set (that is, the proportion of salt mine image pixels in the whole sample image) was analyzed.

As shown in Fig. 1(a), most of the coverage ratios of pixels in salt mine are between 0.0 and 0.1, however, there are 1,562 sample images without salt, which indicates that the data categories in the training set were unbalanced. The unbalance of data categories will cause the designed recognition model to prefer fitting the image data with more categories and ignore the image data with fewer categories, which adversely influences the prediction ability of the designed model. The depth distribution of the salt mine location, that is the imaged depth in the sample images are also analyzed in Fig. 1(b), which indicates that the depth distribution in the training set is basically the same as that in the testing set, and the generalization ability of the designed model could be guaranteed.
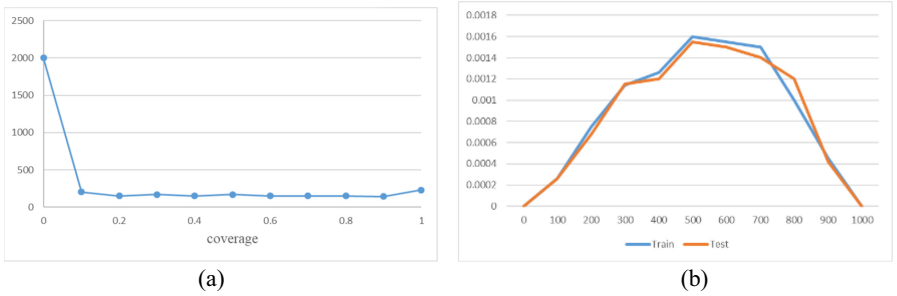

(a)   (b)

Fig. 1. (a) Analysis of salt mine coverage rate of samples in the training set, and (b) the depth distribution of the salt mine location.

Since the original size of the sample image is 101*101, while the size of the commonly used input image in CNN training is 128*128 or 256*256, the bilinear interpolation mode of resize contained in OpenCV is used to adjust the image size. In addition, to make the characteristic values between different dimensions have some comparability in values, and make the network in the designed recognition model has a better learning capability for faster achieving the local optimal, normalization is executed on the pixel values of the sample images.

As the above exploratory data analysis, the unbalance of data categories in the training set will make the designed recognition model be easy to be overfitting on the training set, which adversely influences the prediction ability of the designed model. Therefore, to increase the size of the image data to further enhance the generalization capability of the designed model, the method of data augmentation, e.g., mirror flip, rotation, translation and other basic image operations, is employed to perform adjustments on sample images to obtain new image data [5].

3   Model Design and Optimization

In this work, since all the images in the employed data set are grey-scale images, U-Net [6], as one of the representative image segmentation model with a great advantage on segmenting gray-scale images, is used for designing a salt mine recognition model. A U-Net model consists of an Encoder and a Decoder. To further reduce the overfitting degree

and ensure the generalization capability of the designed model, ResNet34-pretrained is used as the Encoder while removing the first MaxPooling down-sampling layer of ResNet34 to obtain a larger prediction feature map. The modified ResNet34 is shown in Fig. 2. Based on the designed U-Net enabled basic model, the following attempts are conducted to further optimizing the model according to the data characteristics.
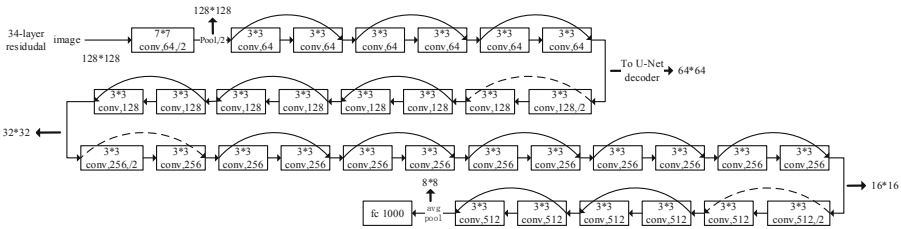

Fig. 2. The modified ResNet34.

Generally, the semantic information of the feature map in the last layer of CNN network is more sensitive to the information of object categories, therefore, many object classification networks always add a full connection layer to the last layer of CNN network for classification and prediction. However, the semantic information of the last layer is very insensitive to these information, e.g., object attitude, background light, object joints and position, etc. Additionally, the features of the last layer are relatively rough in space, which affects the accurate object location. Hence, for fine-grained analysis tasks, such as the image segmentation in this work, direct using the last layer of CNN is often not an optimal choice. Making full use of the feature maps in the shallow layers of CNN network could be an efficient alternative. To this end, as shown in Fig. 3, the hyper-column design is added to the U-Net model to optimize the designed recognition model [7]. Concretely, the feature maps in the upper-sampling layers of multiple stages in Decoder are bilinear interpolated to the same size and then combined in series as the final prediction features for fine-grained target location and object segmentation.
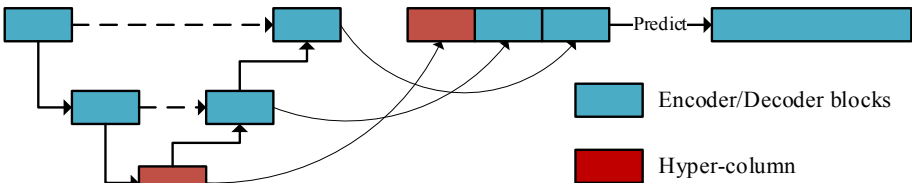

Fig. 3. U-Net model added the hyper-column.

Since the sample images in the employed data set contain images with salt and images without salt, if all the sample images can be classified into images with salt and images without salt in advance, and the segmentation operation could be only executed on the images with salt to improve the recognition efficiency. To this end, an extraction

operation of a binary task branch is executed for the features at the last down-sampling stage in U-Net model [8]. The training principle of the auxiliary segmentation model is shown in Fig. 4, where the weight of the loss function value of the binary classification task is set to 0.05, which is relatively small because it will bring regularization and reduce the degree of overfitting.
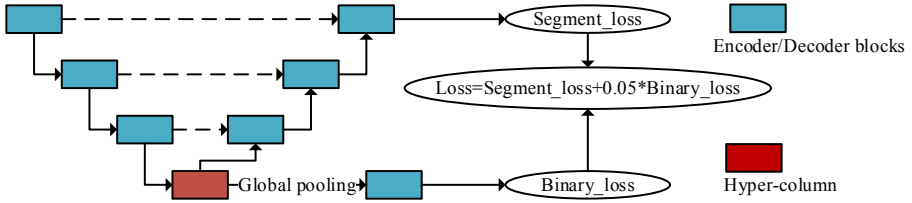

Fig. 4. U-Net model added the binary auxiliary function.

Subsequently, we consider the adjustment and optimization from the network learning to the feature map, and give more weight to the important features while reducing the weight for the unimportant features. In this work, comprehensive considering the advantages of both Spatial Squeeze and Channel Excitation (cSE) and Channel Squeeze and Spatial excitation (sSE) proven in the previous proposals, a combination named scSE is added to the U-Net model to optimize the designed recognition mode [9]. cSE is responsible for the feature reactivation in the channel dimension, while sSE is responsible for the feature reactivation in the spatial dimension, which can comprehensively control the importance of features. As shown in Fig. 5, the locations of the added scSE modules are behind the feature maps of the down-sampling or upper-sampling at each stage to monitor the importance of the features.
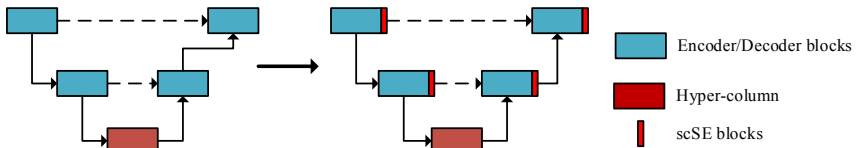

Fig. 5. U-Net model added scSE modules.

Finally, to further enhance the generalization capability of the recognition model and reduce the test error, a depth supervision scheme is added in the U-Net model shown in Fig. 6, where a partition loss function calculation is executed for the feature maps at each upper-layer sampling stage of Decoder and a weighted sum is obtained by multiple loss values [10]. This design considers not only the down-layer characteristics, but also the deep-layer characteristics, which could improve the performance of the model by learning more accurate edge position of the target object in terms of the features in multiple layers.
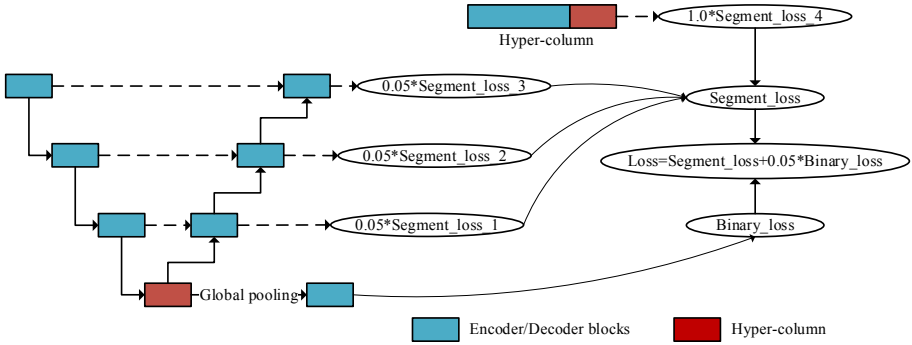
Fig. 6. U-Net model added the depth supervision scheme.

4 Loss Function Design and Optimization

To further improve the generalization capability of the designed recognition model, multiple loss functions are attempted to be adapted in the salt mine image segmentation model.

For ordinary target binary tasks, it is possible to use binary cross entropy (BCE) loss function [11, 12]. In every gradient return, BCE has the same attention to each image category without weight discrimination, so it is very susceptible to the influence of category imbalance. In the every gradient return, the logarithmic loss function has the same attention to each category without weight discrimination, so it is very susceptible to the influence of category imbalance. The exploratory data analysis indicates that the data categories in the training set were unbalanced. Therefore, to overcome this problem, the label generated by the designed image segmentation model is almost a mask that does not contain any location of salt mine. In addition, weighted cross-entropy (WCE) loss function alleviates the problems caused by unbalanced sample categories to some extent, but its disadvantage is that the weights of difficult samples need to be adjusted artificially, which increases the adjustment cost and difficulty [13].

Although WCE needs to artificially adjust the weights of difficult samples, setting a large weight for loss of a difficult sample is necessary to make the model pay more attention to these types of samples. Focal loss function proposed to solve the problem of the serious imbalance between positive and negative samples could make the network model to actively learn these difficult samples [14].

Dice loss function is derived from binary classification by introducing dice coefficient which is essentially an overlap measure of two samples overlap, or similarity of two contour areas [15]. If the value of dice coefficient is increasing while using Dice in the designed image segmentation model, the overlap between the predicted target location and the real target location is increasing as well.

For intersection over union (IOU) acting as a measure of the accuracy of an object in a given set of data, BCE cannot directly optimize the IOU, while Dice Loss can indirectly optimize the IOU. Therefore, under the premise of using BCE, combining Dice as multi loss could achieve multi-angle supervised model training. In addition, since BCE cannot directly optimize the IOU, some researches attempt to avoid using

BCE in the image segmentation model. As a representative method, Lovasz-Softmax Loss function (Lovasz for short) could directly optimize the IOU [16]. In this work, Lovasz is also used in the designed image segmentation model to directly optimize the IOU while effectively reducing loss.

5   Numerical Analysis and Evaluation

In this section, the following numerical analyses are conducted to evaluate the efficiency of achieved proposals in this work.

With the U-Net model and BCE + Dice loss function, the differences in the loss value and accuracy of model training under different data augmentation methods are shown in Fig. 7. From the comparisons, we can clearly see that these data augmentation operations, e.g., padding and horizontal mirror flip, could yield higher benefits, while rotate and direct resize would reduce returns, and a conclusion could be obtained that some data augmentation operations destroy the spatial structure information of the original image and increase the model learning difficulty.
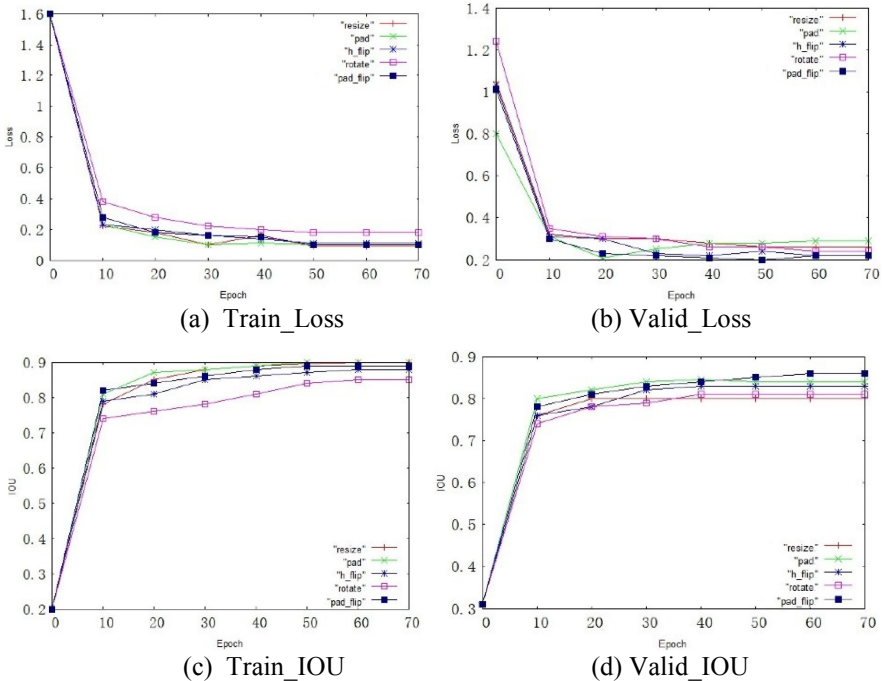

(a) Train_Loss   (b) Valid_Loss

(c) Train_IOU   (d) Valid_IOU

Fig. 7. Differences in the loss value and accuracy of model training under different data augmentation methods.

With the salt mint recognition model designed based on the U-Net, many attempts are conducted to further optimizing the model. The differences in the loss value and accuracy

of different model optimization strategies are shown in Fig. 8. From the comparisons, we can clearly see that the segmentation of small objects and the positioning of object boundaries could be improved by adding hyper-column and scSE modules into the U-Net model, and the overfitting degree of the model on the training set can be effectively reduced by adding auxiliary functions and depth supervision scheme. In addition, the combination of hyper-column and scSE could achieve a greater improvement in model accuracy than that separately adding hyper-column or scSE module.
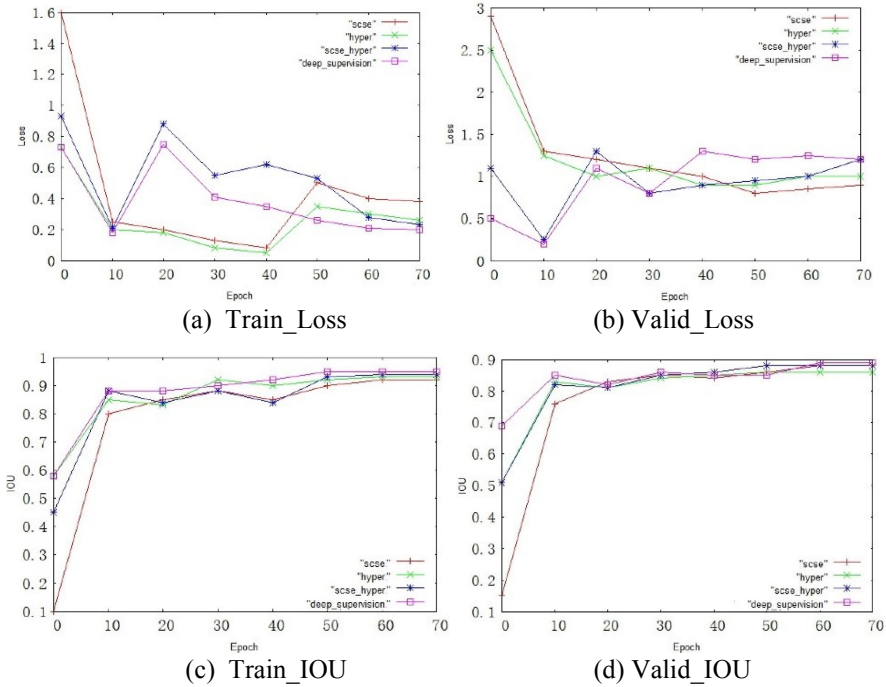

(a) Train_Loss   (b) Valid_Loss

(c) Train_IOU   (d) Valid_IOU

Fig. 8. Differences in the loss value and accuracy of different model optimizations.

In the designed salt mine image segmentation model, multiple loss functions are adapted to further improve the model generalization capability. With U-Net model and padding/horizontal mirror flip enabled data augmentation, the differences in the loss value and accuracy of using different loss functions are shown in Fig. 9. From the comparisons, we can clearly see that BCE is simple and fast to be calculated, but it is easily affected by the imbalance sample category. Lovasz can directly achieve the optimization from the model indicators with high returns but slow efficiency.
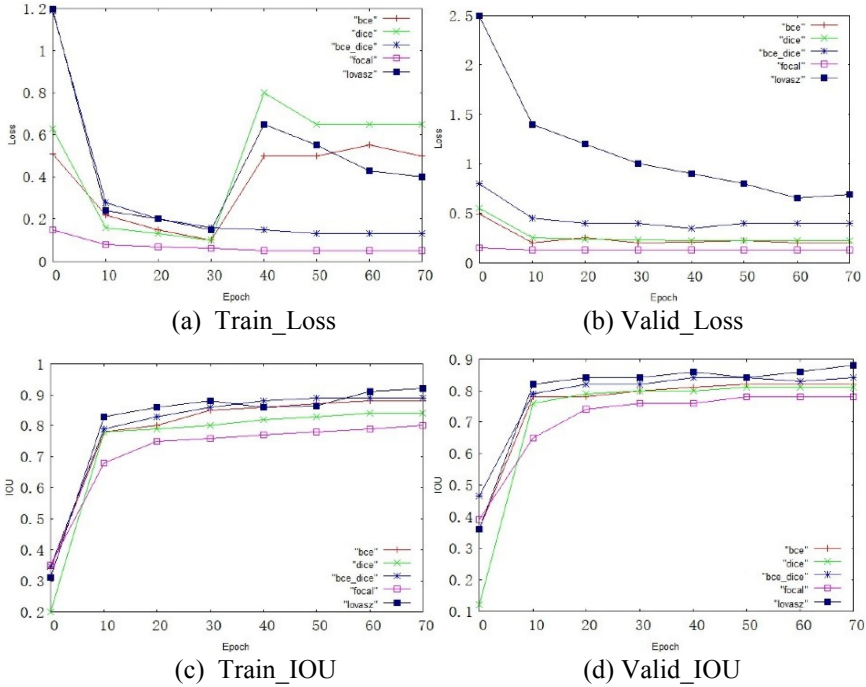
(a) Train_Loss   (b) Valid_Loss

(c) Train_IOU   (d) Valid_IOU

Fig. 9. Differences in the loss value and accuracy of using different loss functions.

6 Conclusion

In this paper, with the in-depth application of deep learning in image segmentation for salt mine recognition, a deep convolutional neural network based image segmentation model is investigated. With exploratory data analysis to mine the characteristics and data processing to increase the size of the image data, a basic recognition model is first designed based on U-Net model integrated modified ResNet34, and many model optimization methods then are used to further optimizing the model according to the data characteristics. Additionally, to further improving the model generalization capacity, multiple loss functions are investigated as well. Finally, numerical analyses are shown to evaluate the efficiencies on loss value and recognition accuracy.

References

1. Liu, F., Lin, G., Shen, C.: CRF learning with CNN features for image segmentation. Pattern Recogn. **48**(10), 2983–2992 (2015)

2. Wang, G., Zuluaga, M.A., Li, W., et al.: DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **41**(7), 1559–1572 (2018)
3. Dolz, J., Gopinath, K., Yuan, J., et al.: HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation. IEEE Trans. Med. Imaging **38**(5), 1116–1126 (2018)
4. Zeng, Y., Jiang, K., Chen, J.: Automatic seismic salt interpretation with deep convolutional neural networks. In: 3rd International Conference on Information System and Data Mining, pp. 16–20 (2019)
5. Fawzi, A., Samulowitz, H., Turaga, D., et al.: Adaptive data augmentation for image classification. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3688–3692 (2016)
6. Ibtehaz, N., Rahman, M.S.: MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. **121**, 74–87 (2020)
7. Zhu, C., Zheng, Y., Luu, K., et al.: Weakly supervised facial analysis with dense hyper-column features. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 25–33 (2016)
8. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: binary or auxiliary supervision. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 389–398 (2018)
9. Roy, A.G., Navab, N., Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Trans. Med. Imaging **38**(2), 540–549 (2018)
10. Pavlakos, G., Zhou, X., Daniilidis, K.: Ordinal depth supervision for 3D human pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition pp. 7307–7316 (2018)
11. Tao, M., Wei, W., Yuan, H., Huang, S.: Version-vector based video data online cloud backup in smart campus. Multimedia Tools Appl. **78**(3), 3435–3456 (2019)
12. Yi-de, M., Qing, L., Zhi-Bai, Q.: Automated image segmentation using improved PCNN model based on cross-entropy. In: International Symposium on Intelligent Multimedia, Video and Speech Processing, pp. 743–746 (2004)
13. Guerrero-Pena, F.A., Fernandez, P.D.M., Ren, T.I., et al.: Multiclass weighted loss for instance segmentation of cluttered cells. In: 25th IEEE International Conference on Image Processing (ICIP), pp. 2451–2455 (2018)
14. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
15. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
16. Berman, M., Rannen Triki, A., Blaschko, M.B.: The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4413–4421 (2018)