Fusion of Deep Learning-based and Spectral Features for Hyperspectral Image Analysis

Evgeny Myasnikov [1]

[1] Samara National Research University, Samara, Russia – mevg@geosamara.ru

Keywords: Hyperspectral Images, Feature Fusion, Classification, Dimensionality reduction, Resnet18.

Abstract

Currently, deep neural networks have become one of the most effective tools in computer vision. However, in the field of hyperspectral remote sensing image analysis, their practical application is limited, as it requires manual labeling of a large amount of data. Since this process is time-consuming and expensive, an attractive option is the use of pre-trained neural networks designed to work with color images. However, to take advantage of hyperspectral images, such neural networks must be equipped with some mechanism to take into account the detailed spectral information contained in such images.
In this work, we propose to combine deep features computed using pre-trained convolutional neural networks (specifically Resnet18) with spectral features of hyperspectral images. The proposed scheme works on the basis of combining the selected type of distances (Euclidean distance, spectral angle, Hellinger divergence) in spectral and embedding spaces with the subsequent synthesis of features in a space of a given dimensionality. The proposed scheme does not require any training, except for the selection of several parameters (spatial window size, dimensionality of the synthesized space, fusion coefficient). Experiments conducted on known hyperspectral scenes (Indian Pines, Salinas, Pavia University, Kennedy Space Center) show the advantages of the proposed approach. The issue of train-test sample splitting is considered.

1. Introduction

Hyperspectral images are an attractive source of remote sensing data. Unlike widely used multispectral data, hyperspectral images have increased spectral resolution. This makes it possible to solve applied remote sensing problems with improved quality.

In recent years, artificial neural networks have been used in image analysis, particularly in the field of remote sensing data. Such networks are increasingly used in the processing of hyperspectral images.

The most widely studied neural network architectures are those based on convolutional neural networks (CNN) [Le Cun et al. 1989]. A one-dimensional CNN-based model for extracting spectral information was proposed [Hu et al. 2015]. However, neural network models that allow taking into account both spectral and spatial features are of greater interest, for example [Chen et al. 2018]. This joint consideration of spectral and spatial information can be implemented by increasing the dimension of convolutional networks. Such a 3D-CNN model was proposed in [Li et al. 2017]. Another method is information fusion, implemented, for example, in the form of two-channel fusion of spectral and spatial information (Two-CNN) [Yang et al. 2017] or multi-channel convolutional neural network (MCCNN) [Chen et al. 2018]. In addition to models that work with original spectral-spatial information, models have been proposed that also allow derived features to be used as input data [Gao et al. 2018].

In addition, methods have been proposed in which feature extraction is performed using ensembles of convolutional neural networks. The networks in ensembles can be of the same type (homogeneous), operating, for example, on random subspaces of the original spectral space [Chen et al. 2019, He and Chen 2021], or of neural networks of various types (heterogeneous) [Nelepa et al. 2021].

Multiscale methods are being explored in which spectral-spatial features are calculated by neural networks for different scales (window sizes) followed by fusion and classification [Safari et al. 2021, Mohan and Venkatesan 2020, Kanthi et al. 2022].

Another class of neural network architectures being explored concerning hyperspectral images is based on graph neural networks (GNN). Such networks consider an image in the form of a graph, where the vertices represent some spatial regions of the image (in a particular case, the image pixels themselves), and the edges connect only adjacent regions. In the field of hyperspectral image analysis, various methods for constructing graphs have been proposed. In [Mou et al. 2020] individual pixels of an image were considered as vertices. [Ren and Zhou 2021] used superpixels (uniform compact groups of pixels). In [Hong et al. 2020] authors used selected image fragments.

In addition to the neural networks described above, other neural network architectures were also used to analyze hyperspectral images. Such architectures are networks with attention mechanisms [Li et al. 2020, Qu et al. 2021], recurrent neural networks (RNN) [Liang et al. 2022], deep belief networks (DBN) [Li et al. 2019], and others.

Unfortunately, training neural networks, especially deep learning neural networks, requires a fairly large labeled sample size. Manual labeling of a large volume of data for training networks requires a lot of time and resources, as ground-based measurements should be carried out.

For this reason, an attractive option may be using neural networks trained on other types of data, such as color or multispectral remote sensing data. In this case, it is necessary to

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-W1-2024-31-2024 | © Author(s) 2024. CC BY 4.0 License.

create some mechanism that allows, firstly, to use of such networks with hyperspectral data, and secondly, to take advantage of the high spectral resolution of hyperspectral images.

This work is devoted to the creation and study of such a mechanism. The work has the following structure. The second section describes the general scheme of the proposed mechanism for applying deep neural networks designed to work with RGB images to hyperspectral images. The section also discusses various methods for taking into account the detailed spectral information contained in hyperspectral images. The third section presents experimental results confirming the effectiveness of the proposed approach. Particular attention is paid to the way of splitting training and test sets in the solution of a classification problem. The paper ends with a conclusion.

2. Methods

This work is based on the feature fusion method described in [Myasnikov 2023]. This method uses an approach based on the fusion of distances between embeddings extracted using deep neural networks with dissimilarities between pixels in hyperspectral space. The obtained distances make it possible to synthesize a new feature space of reduced dimensionality, in which various applied problems can be solved, including pixel-wise classification.

2.1 General scheme of the method

The general scheme of the proposed approach is shown in Figure 1. In this figure, hyperspectral image processing is performed in two independent branches. In the left branch, after selecting the channels corresponding to the natural colors of the input hyperspectral image, feature extraction is performed using the deep neural network (resnet18 in the figure). Such features are extracted for the spatial neighborhoods of each of the analyzed image pixels. After this, the Euclidean distance $d(e_i, e_j)$ can be calculated between any two embeddings e_i и e_j corresponding to the neighborhoods of pixels i and j.

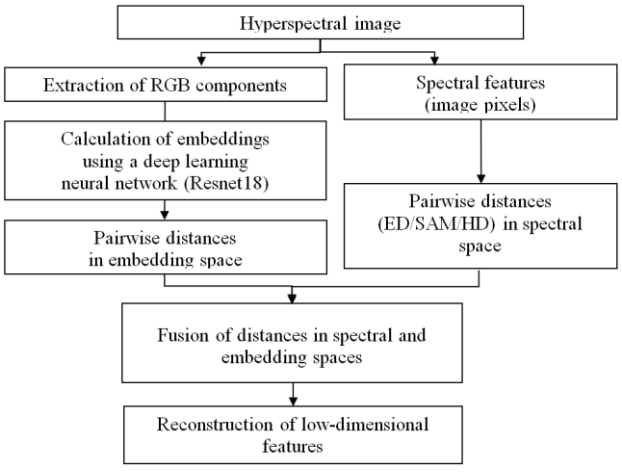


Figure 1. General scheme of the method.

In the right branch of the scheme, hyperspectral image pixels are used directly as spectral features. For any two pixels i and j with spectra s_i and s_j, the distance $\delta(s_i, s_j)$ in spectral space can be calculated in different ways. Following [Myasnikov 2023], in this paper, we used Euclidean distance (ED), spectral angle mapper (SAM) [Kruse et al. 1993], and Hellinger divergence (HD) [Hellinger 1909] as distances in hyperspectral space.

The combined distance in the synthesized Euclidean space Y between pixels i and j can be calculated by the formula:

$$d_{ij} = \left((1-\alpha)^2 d^2(e_i, e_j) + \alpha^2 \delta^2(s_i, s_j)\right)^{1/2}, \quad (1)$$

where α = the fusion coefficient determining the influence of deep or spectral features, $0 \le \alpha \le 1$.

Although the configuration (coordinates) of pixels $Y=\{y_i\}_{i=1..N}$ (N is the number of pixels) in the synthesized space of some given dimensionality m is unknown, these coordinates y_i can be reconstructed by minimizing the error

$$\varepsilon = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\|y_i - y_j\| - d_{ij} \right)^2, \quad (2)$$

Such minimization in this paper was carried out using the gradient descent method:

$$Y(t) = Y(t-1) - \eta \nabla \varepsilon, \quad (3)$$

where η = the gradient descent coefficient (step size),
$\nabla \varepsilon$ = the gradient of the objective function,
t = the iteration number of gradient descent.

Thus, we applied here a nonlinear dimensionality reduction technique, suffering from high computational costs [Myasnikov 2017]. To speed up the calculations, we used stochastic gradient descent with error control based on mini-batches, implemented for GPUs using the CUDA platform.

2.2 Construction of the embeddings using a neural network

As it was mentioned earlier, in this work, we used a fully trained neural network designed for analyzing color (multispectral) remote sensing images to calculate the embeddings. In particular, we used the pre-trained ResNet18 network, which is part of the TorchGeo package [Stewart 2022]. We cut off the classification layers of the network to obtain the embeddings.

To calculate the embedding corresponding to some particular pixel, we superimposed a rectangular window of a predefined size centered on the analyzed pixel. The contents of the window are fed to the input of the neural network, and the output of the cut network is considered as an output embedding. Due to the difference in the window size and the input dimension of the neural network, interpolation is required. Here we relied on recommendations from [Corley et al. 2023] and used bilinear interpolation.

We used ResNet18 SeCo trained using the Seasonal Contrast (SeCo) method [Manas et al. 2021] as a base deep learning model. In our preliminary studies, this model demonstrated higher accuracy compared to the ResNet18 MoCo model trained using the Momentum contrast method [He et al. 2020] (also included in the TorchGeo package [Stewart et al. 2022]).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-W1-2024-31-2024 | © Author(s) 2024. CC BY 4.0 License.

3. Experiments

The main goal of the experiments was to test the feasibility of using neural networks trained using color images to analyze hyperspectral data using the proposed approach to spectral feature fusion.

3.1 Datasets

For the study presented here, we used several well-known hyperspectral image scenes [Baumgardner 2015, Grupo de Inteligencia Computacional, 2021] equipped with groundtruth classification. A brief description of the images is given in Table 1. For brevity, we do not present here the images themselves, true classification masks, and class composition, since the reader can easily find this information using the references above.

Scene	Sensor	Size (HxW)	Bands	Spatial resolut.	Classes
Indian pines	AVIRIS	145x145	220 (200)	20 m.	16
Salinas	AVIRIS	512x217	224 (204)	3.7 m.	16
Pavia University	ROSIS	610x610	103	1.3 m.	9
Kennedy Space Center	AVIRIS	512x614	224 (176)	18 m.	13

Table 1. Test hyperspectral scenes

3.2 Quality assessment methodology

To evaluate the features synthesized using the proposed approach, we solve the most common applied problem, i.e. the problem of pixel-wise multiclass classification. Intuitively, a higher quality of the solution to this problem indicates a higher quality of synthesized features.

To assess the quality of pixel-wise classification, we use the overall classification accuracy (*Acc*), defined as the proportion of correctly classified pixels in the test set.

We use the nearest neighbor (NN) classifier. This choice is due to the following considerations:
- this classifier has no parameters, which eliminates possible bias due to suboptimal classifier parameters for certain scenarios;
- this classifier does not require training, which means there is no need to train or configure any parameters of the entire processing chain, except the size of the spatial window w for the neural network, the fusion coefficient α in expression (1), and the dimensionality of the synthesized space m.

3.3 Sample split

To solve the problem of pixel-wise classification, it is necessary to form training and test samples in one way or another. The data sets described above do not come with standard splits, so random splits are most often used. We propose that this approach may give overestimated quality indicators for methods that use spatial information in addition to spectral information, which is especially true for neural network methods.

For this reason, in this paper, we used the following partitioning method. We took each groundtruth region belonging to one class, determined the most elongated axis (horizontal or vertical), and divided the region into two equal parts along this axis. Examples of the resulting partitions are shown in Figure 2.

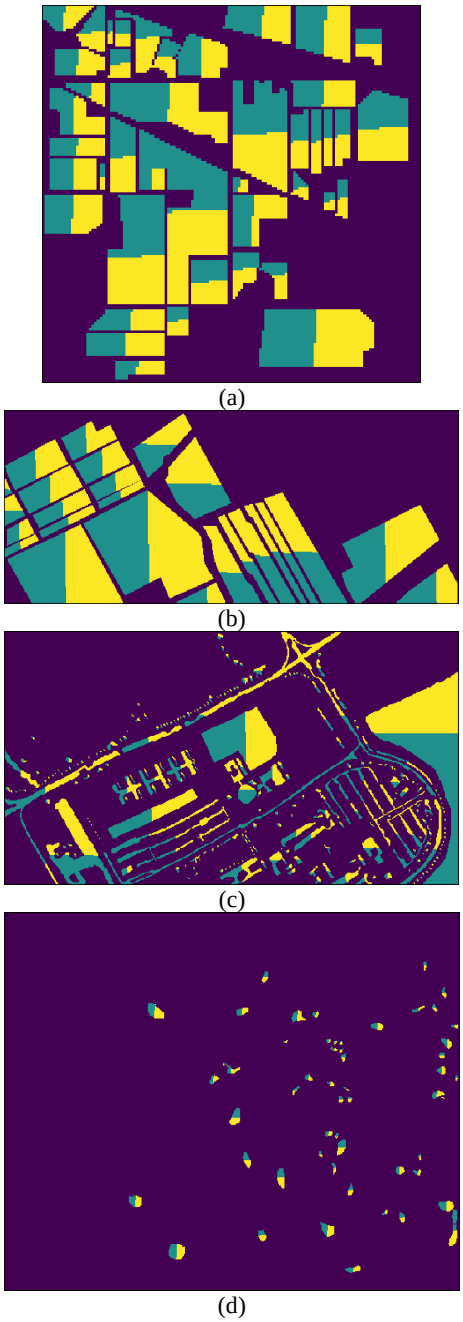


Figure 2. Examples of partitioning into training and test samples for the Indian Pines (a), Salinas (b), Pavia University (c) and Kennedy Space Center (d) scenes.
Images are scaled; images (b, c) are rotated.

One of the resulting halves was used for training, and the other one was used for testing. Then the training and test samples were swapped, and the results were averaged.

3.4 Selection of the window size and dimensionality of the synthesized space

The experiments were divided into several stages. In the first stage, we investigated how the size of the spatial window used in extracting deep features affects the classification quality. We varied the window size from 15 to 33 and measured the classification quality. For each window size, we extracted deep features and combined them with spectral features according to the proposed scheme.

To estimate the difference in hyperspectral space, we used Euclidean distance (ED), spectral angle mapper (SAM), and Hellinger divergence (HD). For each of the spectral dissimilarity measures, the values of the fusion coefficient α that were close to optimal were used (see subsection 3.5). The output dimension for the synthesized space was chosen as m=20. Figure 3(a) shows the experimental results for the Indian Pines scene.

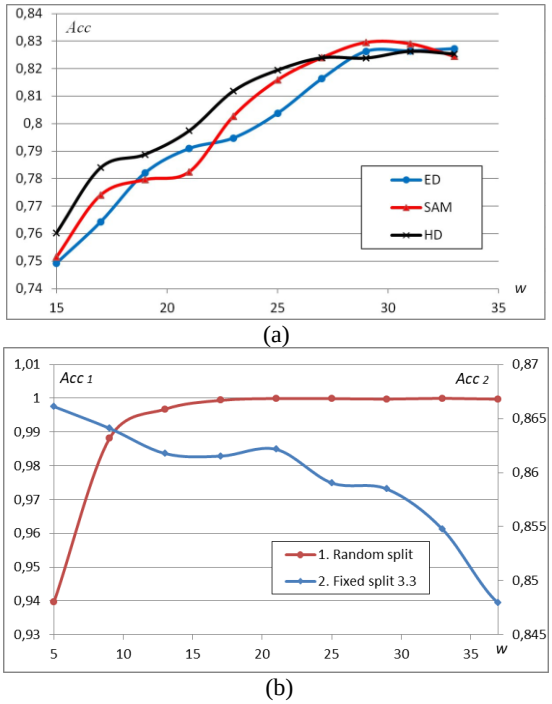


Figure 3. Dependence of classification accuracy *Acc* on the size of the spatial window *w* used for the extraction of deep features for the Indian Pines scene (a) and for the Salinas scene (b).

As can be seen in the figure, the classification quality *Acc* increases with the window size *w* for all spectral dissimilarity measures that is expected. Values *w*=29...31 seem optimal in the selected range. Similar graphs were obtained for Pavia University and Kennedy Space Center.

It is interesting to note that for the Salinas image, the classification quality increased with the window size (Figure 3(b), red line "1") when we used the random split. When we used the fixed partition (subsection 3.3), on the contrary, increasing the window size led to a decrease in classification accuracy (Figure 3(b), blue line "2"). For this reason, a window size of 5 was further used for the Salinas image.

The dependence of classification accuracy *Acc* on dimensionality *m* of the synthesized space is shown in Figure 4.

As can be seen, the quality of classification increases with increasing dimensionality for all the hyperspectral scenes. The greatest changes occur in the range of dimensions *m*=3…20, and after *m*=25 the classification accuracy practically does not change.

Therefore, further in the paper we report the results for dimensions *m*=3..15 in graphical form (so that the graphs do not mix). Also, we present the stable results for *m*=30 in text form.

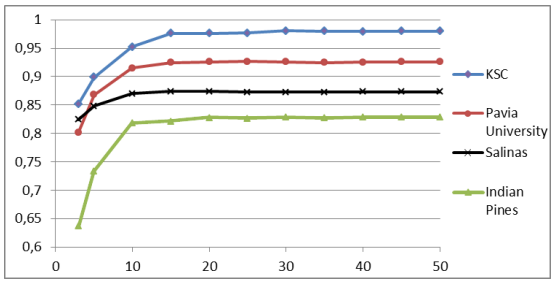


Figure 4. Dependence of classification accuracy *Acc* on the dimensionality *m* of the synthesized space for different hyperspectral scenes.

3.5 Selection of the fusion coefficient

In the second stage of experiments, we investigated the influence of the fusion coefficient α of deep and spectral features on the quality of classification. Here, for four selected scenes from Table 1, we varied the fusion coefficient α from 0 to 1 with a step of 0.1 and measured the classification quality for different dimensions *m* of the synthesized space. The results of the experimental study are shown in Figures 5-8.

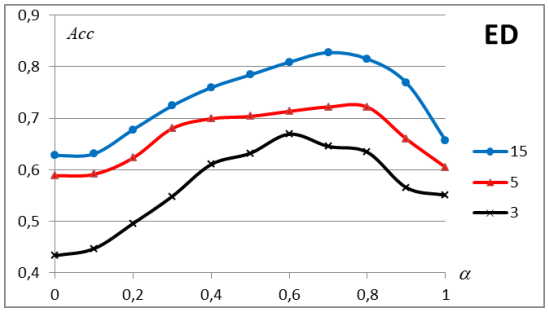
Each figure corresponds to a different hyperspectral image, and each individual part of a figure (a-c) corresponds to a particular measure of spectral dissimilarity (ED, SAM, or HD). Each part of a figure contains three curves corresponding to three different dimensionalities of the synthesized space. For example, in Figure 5, the dimensionality *m* takes the values 3, 5, and 15. The horizontal axis in each figure shows the fusion coefficient α, and the vertical axis shows the classification accuracy *Acc*. Let's take a closer look at the results.

For the Indian Pines image, the classification performance using purely spectral features for each dimension is slightly higher than for deep features. For example, for dimensionality *m*=30 (not shown in the figure), the classification accuracy using deep features was 63.5% versus 64.7%, 65.8%, and 65.8% for ED, SAM, and HD, respectively.

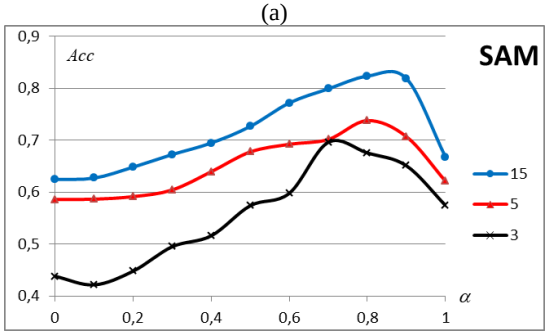
The fusion according to the proposed scheme led to the expected significant improvement in the classification quality. The best-achieved values for dimensionality *m*=30 were 82.8%, 83.5%, and 83.2% for ED, SAM, and HD, respectively, which is 17-18% higher than the results obtained using only spectral features.

The optimal values of the fusion coefficient α were different for different spectral dissimilarity measures. A more detailed study with a step α of 0.01 for *m*=10..30 made it possible to identify the following ranges: for ED $\alpha_{opt} \in [0.77, 0.79]$, for SAM $\alpha_{opt} \in [0.8, 0.83]$, for HD $\alpha_{opt} \in [0.92, 0.94]$. Due to space limitations, we do not include the corresponding graphs.

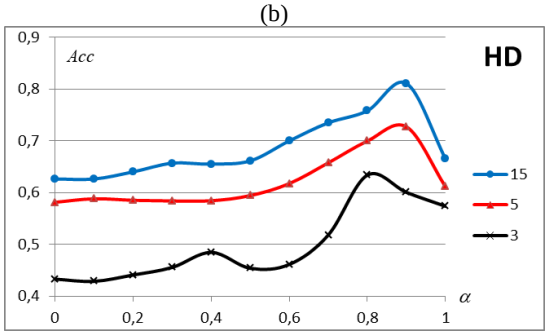
This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-W1-2024-31-2024 | © Author(s) 2024. CC BY 4.0 License.



(a)



(b)

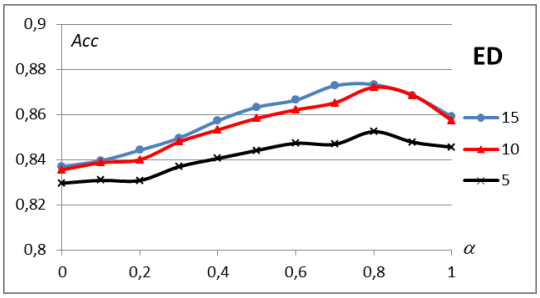


(c)

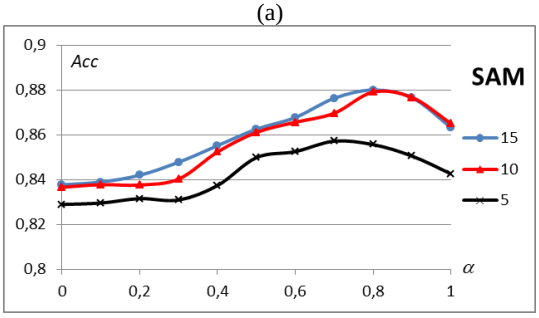
Figure 5. Dependence of the classification accuracy *Acc* on the fusion coefficient α of deep and spectral features for the Indian Pines scene.

The results for Salinas and Pavia University are shown in Figures 6, 7. For the Salinas scene, the best values for dimensionality m=30 obtained using the proposed approach were 87.3%, 88.0%, and 88.2% for ED, SAM and HD, respectively. This is 3.3-4.2% better than the deep features, and 1.2-1.5% better than pure spectral features. For the Pavia University scene, the best values for dimensionality m=30 obtained using the proposed approach were 92.7%, 91.3%, and 90.8% for ED, SAM, and HD, respectively. This is 14.2-16.1% better than deep features and 7.9-9.5% better than pure spectral features. In general, the results for the Salinas and Pavia University scenes are similar to the results for Indian Pines.

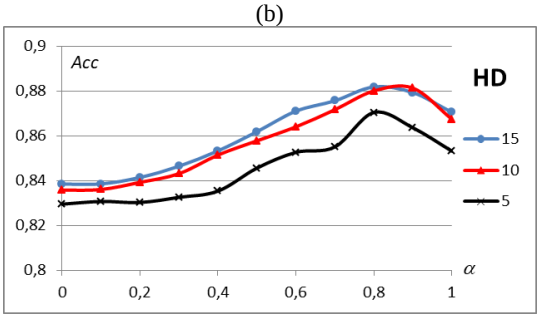
The graphs for the Kennedy Space Center scene are somewhat different (see Figure 8). Here, firstly, the classification quality using deep features (at α=0) turns out to be significantly higher than the quality using purely spectral features (at α=1). So, with m=30 (not shown in the figure), the quality of classification using deep features was 96.5% versus 86.6%, 87.9%, and 88.3% for ED, SAM, and HD, respectively. The best values for dimensionality m=30 achieved by fusing deep and spectral features were 98.0%, 98.2%, and 98.2% for ED, SAM, and HD, respectively, which is 1.5-1.7% higher than the results obtained using deep features only.



(a)



(b)



(c)

Figure 6. Dependence of the classification accuracy *Acc* on the fusion coefficient α of deep and spectral features for the Salinas scene.

The optimal values of the fusion coefficient varied significantly among different spectral dissimilarity measures. A more detailed study with a step α of 0.01 for m=10..30 made it possible to identify the following ranges: for ED $\alpha_{opt}\in$[0.92, 0.96], for SAM $\alpha_{opt}\in$[0.18, 0.2], for HD $\alpha_{opt}\in$[0.28, 0.36]. Due to space limitations, we do not include the corresponding graphs here.

It is worth noting that it is difficult to make a clear choice in favor of one of the spectral dissimilarity measures (ED, SAM, or HD). We conclude that the choice of the optimal spectral dissimilarity measure depends on the data. The difference between the achieved results was not so significant compared to the effect of feature fusion.

3.6 Dependence of results on the splitting method

As a final experiment, we compared the classification accuracy values obtained using deep features for the fixed split used in this paper and the random split in the same 50:50 ratio. The results are presented in Table 2.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-W1-2024-31-2024 | © Author(s) 2024. CC BY 4.0 License.

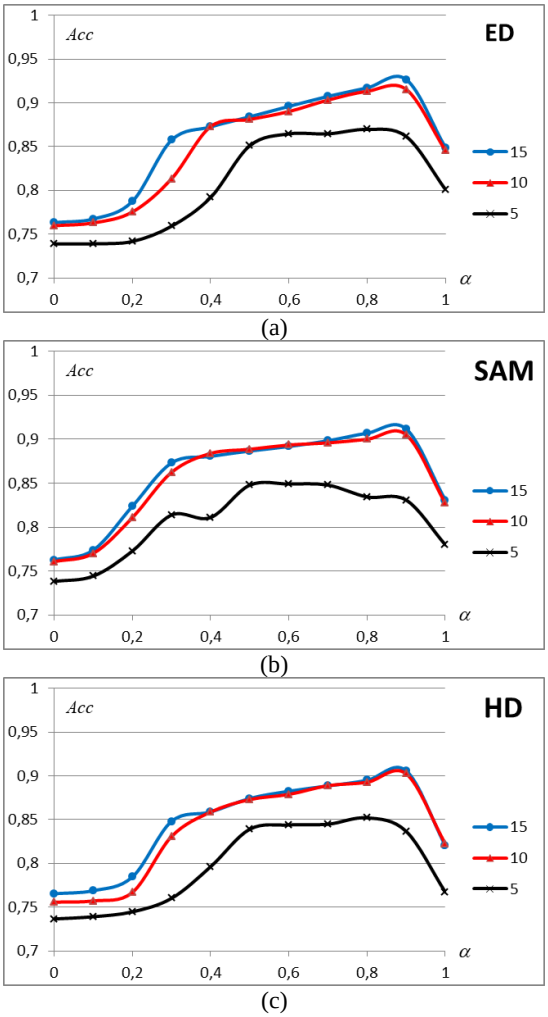


Figure 7. Dependence of the classification accuracy *Acc* on the fusion coefficient α of deep and spectral features for the Pavia University scene.

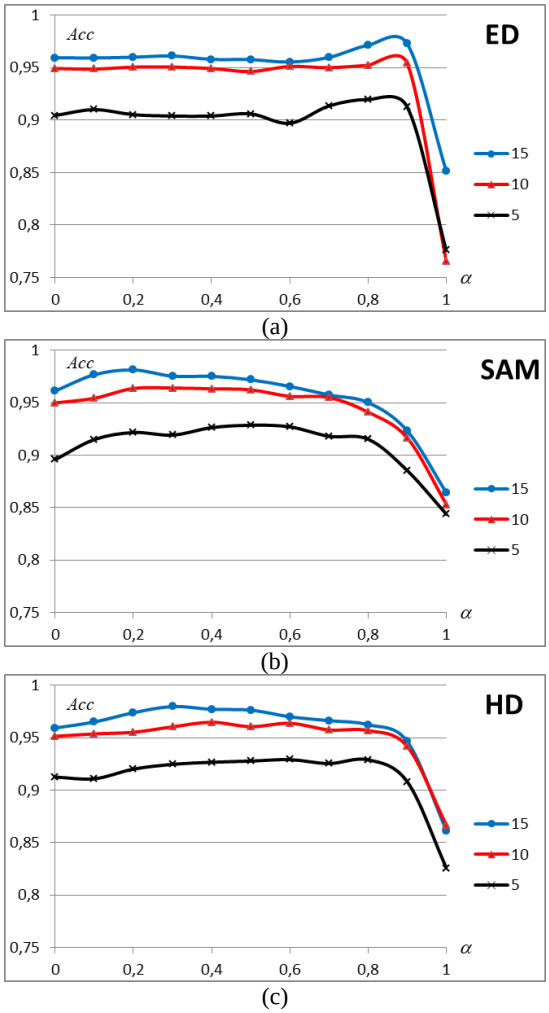


Figure 8. Dependence of the classification accuracy *Acc* on the fusion coefficient α of deep and spectral features for the Kennedy Space Center scene.

The first column of the table shows the results obtained for a random split using only deep features calculated with α=0 and m=30. The second column shows the results for a fixed split, described in subsection 3.3. In the third column, for comparison, we give a range of best accuracy values obtained with the proposed fusion technique for different spectral dissimilarity measures.

As can be seen from Table 2, with random split, the use of only deep features makes it possible to achieve almost error-free classification. When compared to the split used in this paper (split in subsection 3.3), the results seem to be greatly overestimated.

Although the neural network is completely pretrained and is not familiar with the test images, the features it generates for the neighborhood regions are close to each other in the embedding space due to the almost identical (overlapped) spatial context. The use of the split 3.3 reduces the possibility of spatial context overlap. It affects classification accuracy and makes the results less overestimated.

Let's pay attention to the results for the Kennedy Space Center scene. Here, even using a fixed split (subsection 3.3), the classification accuracy remains quite high for the deep features. We believe that the reason for this is the small-sized regions in the ground-truth classification (see Figure 2(d)). Neural network features with a sufficient window size describe the spatial context well and show good results in this case.

Scene	Random split, deep features	Split 3.3 deep features	Split 3.3, proposed technique
Indian pines	99.4	63,5	82.8-83.5
Salinas	99.9	83,9	87.3-88,2
Pavia University	99.9	76,6	90.8-92.7
Kennedy Space Center	100	96,5	98.0-98.2

Table 2. Train-test split comparison for deep and fused features

4. Conclusions

In this paper, we proposed to combine deep features calculated using convolutional neural networks pretrained on color images with spectral features of hyperspectral images. The proposed technique is an example of a feature fusion approach based on

classical methods, although it uses neural networks to extract part of the features. The proposed technique does not require any training or fine-tuning neural networks other than selecting the size of the spatial window, the dimensionality of the synthesized space, and the fusion coefficient.

We conducted a study on four known hyperspectral scenes and determined the parameters of the proposed technique. Using the fusion according to the proposed technique, we obtained a notable increase in the classification quality for all the test images.

In addition, we showed the importance of the train-test splitting procedure for methods that take into account spatial context, since random splitting can produce overestimated results.

Acknowledgments

The research was financially supported by the Ministry of Education and Science of Russia (Agreement No. 075-15-2024-558).

References

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1(4), 541-551. doi.org/10.1162/neco.1989.1.4.541.

Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors* 2015, 258619, 1–12. doi.org/10.1155/2015/258619.

Chen, C., Jiang, F., Yang, C., Rho, S., Shen, W., Liu, S., Liu, Z., 2018, Hyperspectral classification based on spectral–spatial convolutional neural networks. *Engineering Applications of Artificial Intelligence* 68, 165–171. doi.org/10.1016/j.engappai.2017.10.015.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing* 9(1), 67. doi.org/10.3390/rs9010067.

Yang, J., Zhao, Y.Q., Chan, J.C.W., 2017. Learning and transferring deep joint spectral–spatial features for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing* 55(8), 4729–4742. doi: 10.1109/TGRS.2017.2698503.

Chen, C., Zhang, J.J., Zheng, C.H., Yan, Q., Xun, L.N., 2018. Classification of hyperspectral data using a multi-channel convolutional neural network. *Proceedings of the the 14th International Conference on Intelligent Computing (ICIC)*, 81–92.

Gao, Q., Lim, S., Jia, X., 2018. Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning. *Remote Sensing* 10(2), 299. doi.org/10.3390/rs10020299.

Chen, Y., Wang, Y., Gu, Y., He, X., Ghamisi, P., Jia, X., 2019. Deep Learning Ensemble for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(6), 1882-1897. doi.org/10.1109/JSTARS.2019.2915259.

He X., Chen, Y., 2021. Transferring CNN Ensemble for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters* 18(5), 876-880. doi.org/10.1109/LGRS.2020.2988494.

Nalepa, J., Myller, M., Tulczyjew, L., Kawulok, M., 2021. Deep Ensembles for Hyperspectral Image Data Classification and Unmixing. *Remote Sensing* 13(20), 4133. doi.org/10.3390/rs13204133.

Safari, K., Prasad, S., Labate, D., 2021. A Multiscale Deep Learning Approach for High-Resolution Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters* 18(1), 167-171. doi.org/10.1109/LGRS.2020.2966987.

Mohan, A., Venkatesan, M., 2020. HybridCNNbased hyperspectral image classification using multiscale spatiospectral features. *Infrared Physics & Technology* 108, doi.org/10.1016/j.infrared.2020.103326.

Kanthi, M., Sarma, T., Chigarapalle, S., 2022. Multi-scale 3D-convolutional neural network for hyperspectral image classification. *Indonesian Journal of Electrical Engineering and Computer Science* 25(1), 307. doi.org/10.11591/ijeecs.v25.i1.pp307-316.

Mou, L., Lu, X., Li, X., Zhu, X.X., 2020. Nonlocal graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 58(12), 8246–8257. doi.org/10.1109/TGRS.2020.2973363.

Ren, S., Zhou, F., 2021. Semi-supervised classification for PolSAR data with multi-scale evolving weighted graph convolutional network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 2911–2927. doi.org/10.1109/JSTARS.2021.3061418.

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., Chanussot, J., 2020. Graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 59(7), 5966–5978. doi.org/ 10.1109/TGRS.2020.3015157.

Li, R., Zheng, S., Duan, C., Yang, Y., Wang, X., 2020. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sensing* 12(3), 582. doi.org/10.3390/rs12030582.

Qu, L., Zhu, X., Zheng, J., Zou, L., 2021. Triple-Attention-Based Parallel Network for Hyperspectral Image Classification. *Remote Sensing* 13(2), 324. doi.org/10.3390/rs13020324.

Liang, L., Zhang, S., Li, J., 2022. Multiscale DenseNet meets with bi-RNN for hyperspectral image classification. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, 5401–5415. doi.org/ 10.1109/JSTARS.2022.3187009.

Li, C., Wang, Y., Zhang, X., Gao, H., Yang, Y., Wang, J., 2019. Deep Belief Network for Spectral–Spatial Classification of Hyperspectral Remote Sensor Data. *Sensors* 19(1), 204. doi.org/10.3390/s19010204.

Myasnikov, E.V., 2017. Fast techniques for nonlinear mapping of hyperspectral data. *Proceedings of SPIE - The International*

Society for Optical Engineering, 10341, 103411D. doi.org/10.1117/12.2268707

Myasnikov, E., 2023. A Framework for Feature Fusion with Application to Hyperspectral Images. *Lecture Notes in Computer Science* 13644, 509–518. doi: 10.1007/978-3-031-37742-6_39.

Stewart, A.J., Robinson, C., Corley, I.A., Ortiz, A., Lavista Ferres, J.M., Banerjee, A., 2022. TorchGeo: deep learning with geospatial data. *Proceedings of the 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, 19, 1–12. doi.org/10.1145/3557915.3560953.

Corley, I., Robinson, C., Dodhia, R., Lavista Ferres, J.M., Najafirad, P., 2023. Revisiting pre-trained remote sensing model benchmarks: resizing and normalization matters. *arXiv*, 2305.13456. doi.org/10.48550/arXiv.2305.13456.

Kruse, F.A., Boardman, J.W., Lefkoff, A.B., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sensing of Environment* 44, 145-163. doi.org/10.1016/0034-4257(93)90013-N.

Hellinger, E., 1909. Neue Begrundung der Theorie Quadratischer Formen von Unendlichvielen Veranderlichen. *Journal fur diereine und angewandte Mathematik* 136, 210–271.

Manas, O., Lacoste, A., Giró-i Nieto, X., Vazquez, D., Rodriguez, A., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9414–9423.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

Baumgardner, M.F., Biehl, L.L., Landgrebe, D.A., 2015. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3. *Purdue University Research Repository*. doi.org/10.4231/R7RX991C.

Grupo de Inteligencia Computacional, 2021. Hyperspectral Remote Sensing Scenes. ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (18 May 2024).