

基于Transformer模型的高光谱图像分类算法研究

赵尚子欣^{1*}, 袁嘉豪¹, 董岩¹, 陈倩^{2#}

¹上海理工大学光电信息与计算机工程学院, 上海

²上海理工大学管理学院, 上海

收稿日期: 2023年12月18日; 录用日期: 2023年12月31日; 发布日期: 2024年1月31日

摘要

高光谱图像(Hyperspectral image, HSI)分类在遥感领域扮演着关键角色。然而, 处理高光谱图像分类任务时, 遇到了光谱相同但物质不同、光谱不同但物质相同的复杂情况。尽管现有基于卷积神经网络(Convolutional Neural Network, CNN)的方法在局部信息处理方面表现出色, 但它们在表示能力上存在一定限制。为了应对这一挑战, 本文提出了一种综合考虑光谱信息和空间信息的Transformer方法(Spatial Spectral Transformer Network, SSTN), 即本文引入了Transformer结构, 在光谱和空间中倾向于捕捉全局信息。通过构建Transformer, 模型将空间光谱特征有机结合。通过规范的实验研究, 本文发现: 在IndianPines和Houston2013数据集的分类任务中, 本文的方法相较于其它Transformer网络表现更为优越, 并在与其他骨干网络的对比中具有显著的改进。

关键词

高光谱图像, 分类, Transformer, 光谱特征

Research on Hyperspectral Image Classification Algorithm Based on Transformer Model

Shangzixin Zhao^{1*}, Jiahao Yuan¹, Yan Dong¹, Qian Chen^{2#}

¹School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

²Business School, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 18th, 2023; accepted: Dec. 31st, 2023; published: Jan. 31st, 2024

*第一作者。

#通讯作者。

Abstract

Hyperspectral image (HSI) classification plays a crucial role in the field of remote sensing. When dealing with HSI classification tasks, there are complex situations where the spectra are the same but the substances are different, and the spectra are different but the substances are the same. Although existing methods based on Convolutional Neural Networks (CNN) perform well in local information processing, they have certain limitations in representation ability. To address this challenge, a Transformer method that comprehensively considers spectral and spatial information (Spatial Spectral Transformer Network, SSTN) has been proposed. To overcome this limitation, this article introduces a Transformer structure that tends to capture global information in both spectrum and space. By constructing a Transformer, the model organically combines spatial spectral features. The rigorous experimental results show that in the classification task of the Indian Pines and Houston2013 datasets, our method performs better than other Transformer networks and shows significant improvement compared to other backbone networks.

Keywords

Hyperspectral Images, Classification, Transformer, Spectral Features

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景与现状

高光谱遥感是遥感领域崭新的技术,通过整合数百个波段的光谱信息,能够捕捉地表特征微妙变化。高光谱技术在多个领域都有广泛应用,而高光谱图像(HSI)分类则是其关键应用之一,广泛用于矿产勘探[1]、农业开发[2]、环境科学[3]等方面。HSI分类研究的主要挑战在于应对高维光谱数据引发的休斯效应[3],其中同一物质可能呈现不同的光谱信息,而相同的光谱信息可能对应不同的物质。近年来,研究者们通过不同的方法探索如何更有效地处理高光谱数据,提高分类的准确性和鲁棒性。

基于CNN的高光谱图像分类方法是最早被广泛研究的方法之一。研究者们通过卷积层、池化层等操作,提取光谱和空间特征,取得了一系列显著成果。Murali等人[4]提出了一种新的用于HSI分类的3D深度特征提取CNN模型,实验结果显著提高了HSI分类的性能。接着,Zhen等人[5]设计了一种用于HSIs的多尺度光谱空间CNN模型,该模型将多个感受野融合特征与不同层次的多尺度空间特征相结合。然后,Hüseyin等人[6]针对当要提取深层空间和光谱特征时,出现梯度特性的损失和由于深度增加而导致的退化等问题,提出了基于3D卷积神经网络的ResNet50分类方法。Koushikey等人[7]提出了一种多级三维卷积神经网络(3-Dimensional Convolutional Neural Network, 3-DCNN),模型将总体准确度和kappa系数提高了2%~5%。

尽管,基于CNN的分类方法取得了突破性进展,但依然存在面临着很多挑战和局限性。例如:CNN主要通过卷积核在图像上进行滑动来获取局部特征,这意味着每个神经元只能感知输入图像的一个相对较小的局部区域。对于一些需要全局上下文信息的任务,如长距离依赖关系的建模,CNN可能不够有效。为了弥补CNN局限性,将Transformer模型[8]应用于HSI分类是一项令人期待的新研究方向,它能够获取长距离依赖关系。另外,这种新颖的应用可能有助于改善HSI分类的性能,特别是在处理光谱信息复

杂、休斯效应显著的情况下。He 等人[9]提出了一种空间光谱 Transformer (Spatial-Spectral Transformer, SST) 模型, 利用类似于 VGGNet [10]的网络结构来提取空间特征。Qing 等[11]通过引入光谱注意机制, 与 Transformer 中的多注意机制相结合, 有效地捕捉了连续光谱关系另外, Hong 等人[12]设计了一种名为光谱 Transformer (Spectral Transformer, SF)的新模型, 它可以从群体邻近波段学习光谱表示信息并构建 Transformer 模型。尽管, 上述基于 Transformer 的分类方法获得了令人感到满意的性能, 但依然面临着光谱相同但物质不同、光谱不同但物质相同的复杂情况, 特征表示能力上有限的问题。因此, 本文提出了一种基于 Transformer 的高光谱图像分类新方法。

2. 研究方法

本文提出的一种基于空谱 Transformer 的高光谱图像分类新方法(Spatial Spectral Transformer Network, SSTN), 其结构如图 1 所示。首先, 整个图像进行相邻空间处理, 即将图像切分为小数据立方体; 接着, 将所切分的数据按一定空间大小进一步划分, 得到新的数据; 然后, 将每一个划分得到的数据立方体经过空间划分之后作为线性映射的输入; 最后, 图像数据进行位置嵌入和 Transformer 模型进行特征提取, 进而分类。

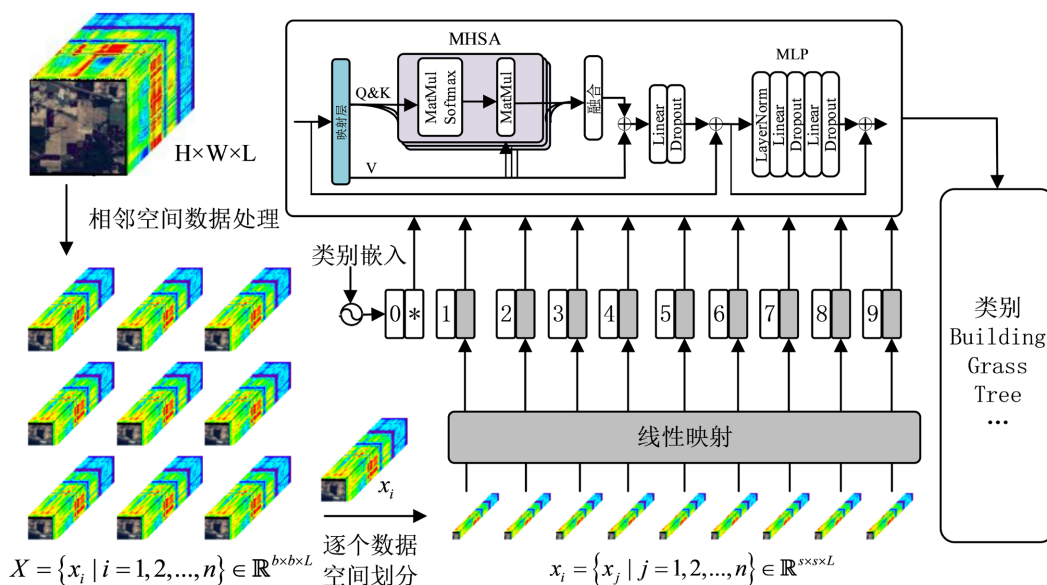


Figure 1. Schematic diagram of SSTN structure
图 1. SSTN 结构示意图

2.1. 相邻空间数据处理

鉴于不同空间包含的地物信息不同, 不同空间位置存在地物相关性。另外, 若将整个图像作为模型的输入数据, 将会给网络带来巨大的计算负担。因此, 考虑到上述问题, 空谱 Transformer 网络首先将输入进行相邻空间处理, 即将其通过逐像素提取获得数据立方体, 其中立方体包含了类别的相邻信息。假设输入的数据大小为 $\mathbb{R}^{H \times W \times L}$, 经过相邻空间数据处理之后得到 n 个 $x_i \in \mathbb{R}^{b \times b \times L}$ 。其中, n 为数据样本总数量。

2.2. 空谱 Transformer 模块

空谱 Transformer 模块即充分考虑了空间信息, 同时还获取了光谱的长距离依赖关系。其结构如图 1 的右半部分。

首先, 为了进一步考虑图像的空间关系, 空谱 Transformer 模块将相邻空间数据处理后得到的数据立方体进行空间 9 等划分, 得到 9 个小数据立方体。此部分可以被表示为:

$$x_i = \{x_j \mid j = 1, 2, \dots, k\} \in \mathbb{R}^{s \times s \times L} \quad (1)$$

上式中, $s \times s$ 表示划分后的数据空间大小, L 表示光谱数量, j 表示划分的个数。

其次, Transformer 模块将划分后得到的数据作为输入的令牌(Tokens), 每个令牌表示为 $[T_1, T_2, \dots, T_z]$, 所有的 tokens 与一个可学习的分类令牌 T_0^{cls} 进行类别嵌入, 线性映射和类别嵌入可被表示为:

$$T_m = [T_0^{cls}, T_1, \dots, T_z] + PE \quad (2)$$

上式中, T_m 表示 Transformer 的输入。

然后, 经过线性映射和类别嵌入的数据作为 MHSA 和 MLP 块的输入。即将数据为 Q 、 K 和 V , Q 和 K 进行多组自注意力特征提取, 最后经过融合与 V 进行乘积。此部分可被表示为:

$$SA(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

$$\text{MHSA} = \text{Concat}(SA_1, SA_2, \dots, SA_h)W \quad (4)$$

上式中, d_k 表示 K 的维度, h 表示多头自注意力的组数。Concat(\cdot) 表示级联函数。另外, 得到的 MHSA 特征被送至 MLP 层进行进一步的高级语义特征的提取。

最后, 经过空谱 Transformer 模块进行特征提取所得的高级特征被送到 Softmax 分类器中进行分类。

3. 实验结果与分析

3.1. 数据集描述

为了验证所提出的 SSTN 模型的有效性, 本文选择了 2 种数据集进行一系列的实验, 分别为 Indian Pines 和 Houston2013 数据集。

Indian Pines 数据集包含了来自不同地物类别的 16 个类别, 其中主要涵盖了农业、森林和植被等地物, 是一副 145×145 像素的图像, 涵盖了 224 个光谱波段的信息。除去吸水波段和低信噪比波段外, 使用了 200 个波段。

Houston2013 数据集是由高光谱图像分析团队和 NCALM (National Center for Airborne Laser Mapping) 在休斯敦大学的校园和附近城区采集的, 图像空间大小为 349×1905 像素, 包含 144 个光谱波段。另外, 该数据集包含了 15 个不同的土地覆盖类别, 这些类别可以包括城市建筑、道路、植被、水体等。

Table 1. Category names and number of data sample divisions for all datasets

表 1. 所有数据集的类别名称和数据样本划分数量

Indian Pines				Houston2013			
Class	Class name	Training	Test	Class	Class name	Training	Test
1	Alfalfa	4	42	1	Healthy grass	125	1126
2	Corn-notill	142	1286	2	Stressed grass	125	1129
3	Corn-mintill	82	748	3	Synthetic grass	69	628

续表

4	Corn	23	214	4	Trees	124	1120
5	Grass-pasture	48	435	5	Soil	124	1118
6	Grass-trees	72	658	6	Water	32	293
7	Grass-pasture-mowed	3	25	7	Residential	126	1142
8	Hay-windrowed	47	431	8	Commercial	124	1120
9	Oats	3	17	9	Road	125	1127
10	Soybean-notill	97	875	10	Highway	122	1105
11	Soybean-mintill	245	2210	11	Railway	123	1112
12	Soybean-clean	59	534	12	Parking Lot 1	123	1110
13	Wheat	20	185	13	Parking Lot 2	46	423
14	Woods	126	1139	14	Tennis Court	42	386
15	Bldg-Grass-Tree-Drivers	38	348	15	Running Track	65	595
16	Stone-Steel-Towers	9	84		/	/	/
/	Total	1018	9231	/	Total	1495	13,534

3.2. 实验设置

本文提出的方法是在Pytorch框架上实现的，设备配备有Intel(R) Core (TM) i9-9900K CPU、NVIDIA GeForce RTX 3090 GPU 和 64 G 随机存取内存。学习率大小为 0.0005，训练轮次为 300。

3.3. 实验定量分析

对于表 1~3 所示的所有方法分类的定量指标，本文方法在 OA、AA 和 Kapp 评价指标上表现优于其他方法。在 Indian Pines 数据集上，SSTN 的 OA 值分别高出 CDCNN、FDSSC、SSRN、ViT、SSFTT 方法 22.74%、7.62%、3.17%、4.18% 和 0.95%。在 Houston2013 数据集，SSTN 的 OA 值分别高出 CDCNN、FDSSC、SSRN、ViT、SSFTT 方法 21.10%、1.91%、16.52%、15.38% 和 4.74%。这充分说明了本文所提出方法的优越性和优异的分类精度。

Table 2. Classification results of all methods on the Indian Pines dataset

表 2. 所有方法在 Indian Pines 数据集上的分类结果

方法	CDCNN	FDSSC	SSRN	ViT	SSFTT	SSTN
1	29.69	73.06	75.46	92.25	90.26	99.89
2	67.31	83.31	90.54	88.85	92.61	88.38
3	57.68	89.05	96.9	71.68	95.52	97.01
4	64.05	75.01	93.25	54.85	89.24	100
5	94.35	87.91	98.45	72.21	96.81	100
6	91.12	92.89	98.83	93.44	99.07	97.07
7	63.15	39.89	84.1	79.15	79.46	92.48
8	88.77	96.17	98.25	99.79	99	99.67
9	62.91	67.67	97.5	50.42	97.8	100

续表

10	65.52	79.18	80.92	90.74	91.97	89.76
11	73.17	90.66	92.91	96.97	95.04	99.22
12	46.74	88.26	95.61	76.74	90.6	98.39
13	90.21	94.46	99.7	93.22	99.7	95.91
14	90.08	96.2	95.68	95.01	97	93.11
15	76.42	66.46	92.14	86.91	88.02	96.88
16	92.82	84.23	96.46	84.84	96.94	94.27
OA	72.53	87.65	92.1	91.09	94.32	95.27
AA	72.12	72.72	92.92	82.94	93.7	96.37
Kappa	68.62	85.93	91.04	89.74	93.54	94.63

Table 3. Classification results of all methods on the Houston 2013 dataset**表 3.** 所有方法在 Houston2013 数据集上的分类结果

方法	CDCNN	FDSSC	SSRN	ViT	SSFTT	SSTN
1	76.67	88.46	86.4	74.26	86.51	97.79
2	81.45	94.5	85.87	78.3	99.16	87.43
3	56.83	99.69	96.7	73.15	99.74	100
4	86.44	99.27	77.96	78.42	93.68	98.11
5	94.93	96.22	92.65	92.45	94.98	99.81
6	84.72	99.73	26.37	76.83	83.5	93.86
7	66.42	84.91	59.97	54.16	66.8	84.23
8	77.94	93.99	63.66	92.78	58.63	66.01
9	64.8	78.51	62.36	68.47	90.64	90.95
10	46.08	71.42	74.44	65	90.61	96.52
11	52.29	86.84	83.44	81.96	75.94	68.71
12	57.07	83.16	81.12	74.52	57.19	90.49
13	80.01	84.14	15.62	73.66	94.99	96.18
14	76.03	97.28	83.11	74.72	100	100
15	82.85	96.74	40.32	89.01	100	100
OA	68.68	87.87	73.26	74.4	85.04	89.78
AA	72.3	90.32	68.67	77.07	86.15	91.34
Kappa	66.15	86.89	71.03	72.24	83.84	88.97

3.4. 可视化分析

所有方法在 Indian Pines 和 Houston2013 数据集上的分类可视化图如图 2 和图 3 所示。在图 2 中，不同的物体之间，所有的分类可视化图类别边缘周围仍存在一些噪声点。与其他方法相比，本文所提出的方法能够形成一个相当清晰的分类图。在图 3 中，通过观察，网球场、停车场和跑道等大型建筑的显著

存在, 可以发现所提出的方法可以有效地保存这种大型结构。

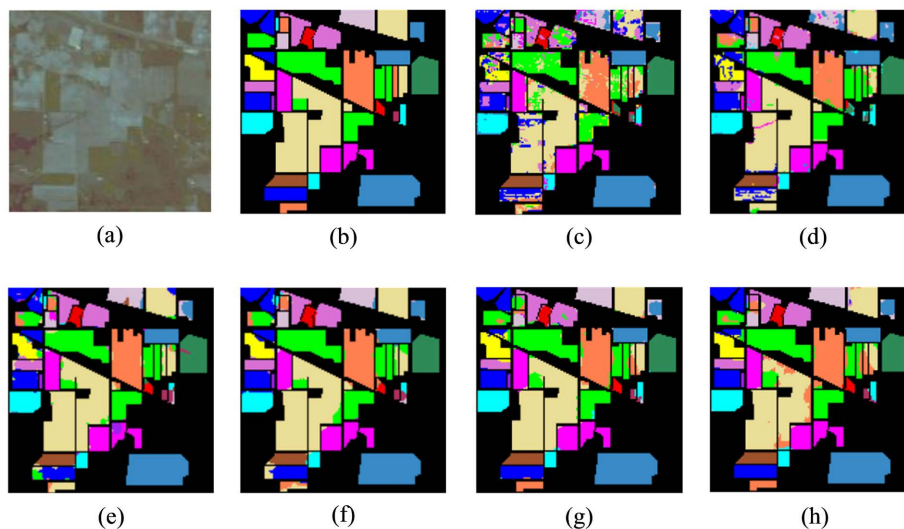


Figure 2. Visualization of classification of all methods on the Indian Pines dataset. (a) Truth-land cover map, (b) false color map, (c) CDCNN, (d) FDSSC, (e) SSRN, (f) ViT, (g) SSFTT, (h) SSTN

图 2. 所有方法在 Indian Pines 数据集上的分类可视化图。(a) 地物真实图, (b) 假彩色图, (c) CDCNN, (d) FDSSC, (e) SSRN, (f) ViT, (g) SSFTT, (h) SSTN

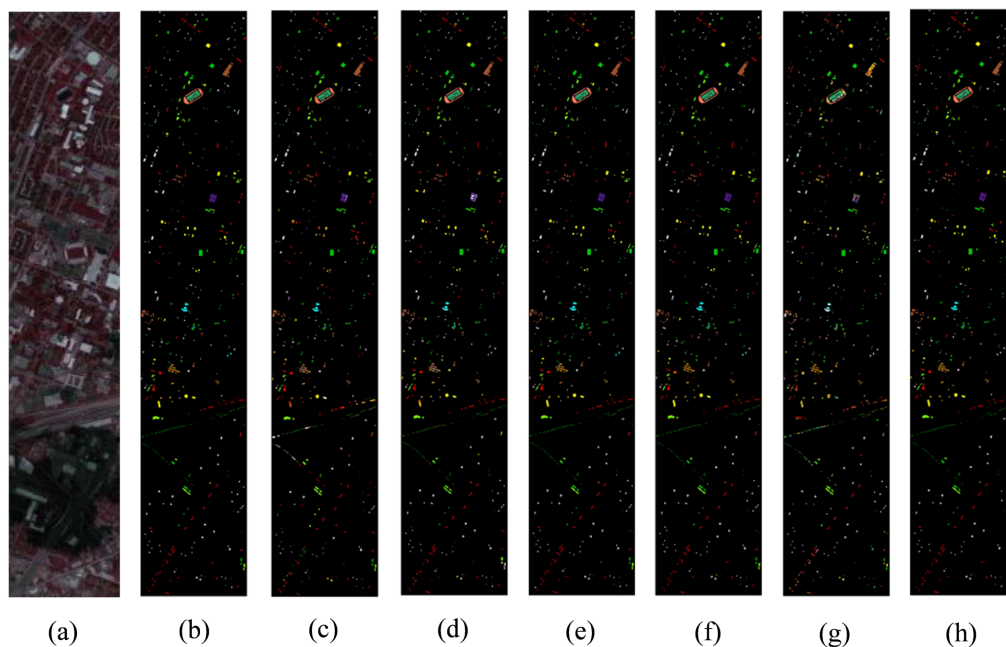


Figure 3. Visualization of classification of all methods on the Houston2013 dataset. (a) Truth-land cover map, (b) false color map, (c) CDCNN, (d) FDSSC, (e) SSRN, (f) ViT, (g) SSFTT, (h) SSTN

图 3. 所有方法在 Houston2013 数据集上的分类可视化图。(a) 地物真实图, (b) 假彩色图, (c) CDCNN, (d) FDSSC, (e) SSRN, (f) ViT, (g) SSFTT, (h) SSTN

4. 总结

本文提出了一种综合考虑光谱信息和空间信息的 Transformer 方法(SSTN)。为了克服现有 CNN 在局

部信息处理方面的局限性, 首先引入了 Transformer 结构, 在光谱和空间中倾向于捕捉全局信息。通过构建 Transformer, 模型将空间光谱特征有机结合。接着, 为了验证 SSTN 方法的有效性, 本文还进行了一系列实验, 包括定量分析和可视化分析。实验结果表明, 本文的方法相较于其他 Transformer 网络表现更为优越, 并在与其他骨干网络的对比中具有显著的改进。

在未来, 这项工作可以扩展, 包括混合深度光谱空间特征提取模型, 如图神经网络, 以提高分类精度。另外, 我们可以进一步提高光谱空间特征学习能力, 寻求速度和精度之间的平衡, 这将是卫星和无人飞行器(UAV)平台进行实时分类的基础。

基金项目

2023 年教育部产学合作协同育人项目(230714421107183); 2023 年“上海高校青年教师培养资助计划”; 2023 年上海市“教师专业发展工程”高校教师培养计划; 上海市青年科技启明星计划(23YF1429700); 国家自然科学基金青年项目(62303325)。

参考文献

- [1] Carrino, T., Crósta, A., Toledo, C. and Silva, A. (2018) Hyperspectral Remotesensing Applied to Mineral Exploration in Southern Peru: A Multiple Dataintegration Approach in the Chapi Chiara Gold Prospect. *International Journal of Applied Earth Observation and Geoinformation*, **64**, 287-300. <https://doi.org/10.1016/j.jag.2017.05.004>
- [2] Camino, C., González-Dugo, V., Hernández, P., Sillero, J. and Tejada, P. (2018) Improved Nitrogen Retrievals with Airborne-Derived Fluorescence and Planttraits Quantified from VNIR-SWIR Hyperspectral Imagery in the Context of Precision Agriculture. *International Journal of Applied Earth Observation and Geoinformation*, **70**, 105-117. <https://doi.org/10.1016/j.jag.2018.04.013>
- [3] Stuart, M.B., McGonigle, A.J.S. and Willmott, J.R. (2019) Hyperspectral Imaging in Environmental Monitoring: A Review of Recent Developments and Technological Advances in Compact Field Deployable Systems. *Sensors*, **19**, 3071. <https://doi.org/10.3390/s19143071>
- [4] Kanthi, M., Sarma, T.H. and Bindu, C.S. (2020) A 3D-Deep CNN Based Feature Extraction and Hyperspectral Image Classification. 2020 *IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, Ahmedabad, 1-4 December 2020, 229-232. <https://doi.org/10.1109/InGARSS48198.2020.9358920>
- [5] Xu, Z., Yu, H., Zheng, K., Gao, L. and Song, M. (2021) A Novel Classification Framework for Hyperspectral Image Classification Based on Multiscale Spectral-Spatial Convolutional Network. 2021 *11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, 24-26 March 2021, 1-5. <https://doi.org/10.1109/WHISPERS52202.2021.9483998>
- [6] Firat, H. and Hanbay, D. (2021) Classification of Hyperspectral Images Using 3D CNN Based ResNet50. 2021 *29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, 9-11 June 2021, 1-4. <https://doi.org/10.1109/SIU53274.2021.9477899>
- [7] Chhapariya, K., Buddhiraju, K.M. and Kumar, A. (2022) Spectral-Spatial Classification of Hyperspectral Images with Multi-Level Cnn. 2022 *12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Rome, 13-16 September 2022, 1-5. <https://doi.org/10.1109/WHISPERS56178.2022.9955063>
- [8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognitionat Scale. arXiv: 2010.11929, 2020.
- [9] He, X., Chen, Y. and Lin, Z. (2021) Spatial-Spectral Transformer for Hyperspectral Image Classification. *Remote Sensing*, **13**, 498. <https://doi.org/10.3390/rs13030498>
- [10] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks Forlarge-Scale Image Recognition. arXiv: 1409.1556, 2014.
- [11] Qing, Y., Liu, W., Feng, L. and Gao, W. (2021) Improved Transformer Net for Hyperspectral Image Classification. *Remote Sensing*, **13**, 2216. <https://doi.org/10.3390/rs13112216>
- [12] Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., et al. (2022) SpectralFormer: Rethinking Hyperspectral Image Classification With Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-15. <https://doi.org/10.1109/TGRS.2021.3130716>