

КОМПЛЕКСНЫЕ АЛГОРИТМЫ КЛАССИФИКАЦИИ ВИЗУАЛЬНЫХ ОБЪЕКТОВ НА БАЗЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Мостовщиков Д.Н.¹, Могилатов Р.К.², Бабкин О.В.³
Email: Mostovshchikov17163@scientifictext.ru

¹Мостовщиков Дмитрий Николаевич – старший системный архитектор,
Li9, Inc., г. Феникс;

²Могилатов Роман Константинович - технический руководитель,
Python SoftServe Inc., г. Роли;

³Бабкин Олег Вячеславович – старший системный архитектор,
Li9, Inc., г. Феникс,
Соединенные Штаты Америки

Аннотация: проведен анализ современных подходов, применяемых при построении программных и нейросетевых алгоритмов выделения и классификации визуальных объектов в массивах графических данных. Указана необходимость комплексного подхода при организации систем машинного анализа изображения. Предложена комплексная схема машинного анализа, которая включает себе этапы предварительной обработки входных данных (операции восстановления изображения, настройки параметров, сегментации) и работы с визуальными объектами (выделение визуального объекта, определение признаков, классификация). Методика оптимизации алгоритмов машинного анализа базируется на математическом моделировании процессов регистрации и обработки изображения, а также построении целевых функций путем формализации требований, указанных при постановке конкретной задачи по работе с графическими данными. Таким образом, в рамках исследования, проблема оптимизации решается через определения глобальных экстремумов целевых функций, которые являются показателями эффективности выделения и классификации визуальных объектов.

Ключевые слова: графические данные, машинный анализ, визуальный объект, программные алгоритмы, нейросетевые алгоритмы, предварительная обработка изображения, целевые функции.

COMPLEX ALGORITHMS FOR THE CLASSIFICATION OF VISUAL OBJECTS BASED ON MACHINE LEARNING METHODS

Mostovshchikov D.N.¹, Mogylatov R.K.², Babkin O.V.³

¹Mostovshchikov Dmitrii Nikolaevich – Senior Systems Architect
LI9, INC., PHOENIX;

²Mogylatov Roman Kostyantynovich – Technical Leader,
PYTHON SOFTSERVE INC., RALEIGH;

³Babkin Oleg Vyacheslavovich – Senior Systems Architect,
LI9 INC., PHOENIX,
UNITED STATES OF AMERICA

Abstract: the analysis of modern approaches used in the construction of software and neural network algorithms for the selection and classification of visual objects in graphical data arrays is carried out. The need for an integrated approach to the organization of systems for machine image analysis is indicated. A complex scheme of machine analysis is proposed, which includes the stages of preliminary processing of input data (image restoration operations, setting parameters, segmentation) and working with visual objects (highlighting a visual object, determining features, classification). The optimization technique for machine analysis algorithms is based on mathematical modeling of the registration and image processing processes, as well as the construction of target functions by formalizing the requirements specified in the formulation of a specific task for working with graphic data. Thus, within the framework of the study, the optimization problem is solved through the determination of the global extrema of the objective functions, which are indicators of the efficiency of the selection and classification of visual objects.

Keywords: graphic data, machine analysis, visual object, software algorithms, neural network algorithms, image preprocessing, objective functions.

УДК 004.021

Введение

Активное развитие информационных технологий и, в частности, появление парадигмы «больших данных» (Big Data, BD), в рамках которой подразумевается экспоненциальный рост данных при расширении инструментария по глубинному анализу массивов данных на машинном уровне (Data Mining) привело к дисбалансу между уровнем поставленной задачи и доступным вычислительным ресурсом. Значительным образом это касается именно работы с графическими данными и видеоданными. С одной стороны, это связано с информационным объемом указанного формата данных и нетривиальностью задачи выделения и классификации визуальных объектов, в отличие от, например, задачи машинного анализа данных представленных в аудио-, текстовом или численном формате. С другой стороны это связано и с ростом объемов графических данных, что является следствием активного внедрения и расширения функционала средств фото- и видео-регистрации, а также перевода аналоговых изображений в цифровую форму [1, 2]. На сегодняшний

день, аппаратно-программные платформы, которые, в том числе, присутствуют в мобильных устройствах, активно распространяются и совершенствуются, а собранные данные передаются и аккумулируются на информационных ресурсах в рамках сетевых протоколов [3, 4]. Таким образом, задача построения автоматизированных систем, способных в режиме реального времени эффективно обрабатывать и анализировать большие объемы графических данных и видеоданных при условии ограничения на вычислительный ресурс информационных узлов сетевых сервисов [5, 6] на сегодняшний день характеризуется чрезвычайно **высокой актуальностью**.

Анализ **современных исследований и публикаций в профильных изданиях** посвященных данной проблеме показал комплексность задачи обработки больших объемов графических данных. Задача машинного анализа была разделена на следующие группы подзадач:

- предварительная обработка графических данных [13-14], которая включает в себя операции по восстановлению изображения [7-9], настройке параметров [10-12] и сегментации [13, 14];
- работы с визуальными объектами [15-25], которая включает в себя выделение визуального объекта [15-18], выделение его признаков [19, 20], а также его дальнейшую классификацию в соответствии с обучающей выборкой [21, 22].

В области предварительной обработки графических данных были рассмотрены математические модели шума [9, 23], морфологические методы обработки [24, 25], пороговые методы [8] и текстурные дескрипторы [26], а также нейросеть типа «автокодировщик» [27, 28]. В свою очередь, в области работы с визуальными объектами рассмотрены нейросетевые алгоритмы, в частности архитектуры многоярусного автокодировщика [21, 29], сверточных нейросетей [30] и SVM-классификатора [26, 31], а также алгоритмы на основе метода Виолы-Джонса [18, 32].

Комплексность задачи оптимизации работы с графическими данными с целью выделения и классификации визуальных объектов, а также большое количество подходов в этой области на уровне построения математических моделей и программных, а также нейросетевых алгоритмов, подразумевает необходимость построения универсальной методики машинного анализа, что выделяется как **нерешенная часть** вопроса в рамках общего исследования.

Целью работы, таким образом, стала разработка универсальной модели машинного анализа больших объемов графических данных и видеоданных с ограничением на время обработки входных запросов и вычислительный ресурс системы.

Основная часть исследования

Выше были рассмотрены общие принципы организации и оптимизации машинного анализа графических данных, где общая схема включает в себе выбор и определение особенностей цифровой системы регистрации, предварительную обработку входных данных, а также выделение визуальных объектов (рис. 1).

Для построения универсальной математической модели машинного анализа необходимо рассмотреть основные этапы получения и обработки входных графических данных, а также выделения и классификации визуальных объектов. Пусть массив входных данных представляет собой набор $A: \{A_n\}$, где A_n — элементы массива изображений или кадры видеофрагмента. В общем случае, любое изображение A_n ($n \in [1; N]$) может быть представлено как набор двумерных матриц $A_n: \{A_n^i\}$ размерности $X_n^A \times Y_n^A$, где $i \in [1; I]$ и для $\forall i A_n^i \in [0; a_n]$. Размерность $X_n^A \times Y_n^A$ соответствует разрешающей способности изображения n , I — количеству каналов цвета, а, в свою очередь, a_n — динамическому диапазону. Так, например, представление изображения в формате «Truecolor» задает значения $I = 3$ (RGB-схема) или $I = 4$ (RGB-схема с дополнительным альфа-каналом, который определяет уровень прозрачности пикселей) и $a_n = 255$ (выделение восьми бит на каждый канал), что подразумевает 24- или 32-битное кодирование, соответственно. В том случае, когда набор $\{A_n\}$ соответствует набору кадров видеофрагмента, величина N также соотносится с общей продолжительностью видеофрагмента T_{Σ} , для определения кадровой частоты $\frac{T_{\Sigma}}{N}$.

Также следует отметить, что набор $A: \{A_n\}$ содержит графические данные, которые были искажены вследствие несовершенства системы регистрации (наличие шумов, оптических aberrаций, низкая разрешающая способность либо цветовая схема и динамический диапазон подлежат коррекции). Пусть массив входных графических данных без искажений может быть представлен набором $B: \{B_n\}$, причем изображение B_n ($n \in [1; N]$) может быть представлено как набор двумерных матриц $B_n: \{B_n^j\}$ размерности $X_n^B \times Y_n^B$, где $j \in [1; J]$ и для $\forall j B_n^j \in [0; b_n]$, причем в общем случае следует указать, что $X_n^B \neq X_n^A$, $Y_n^B \neq Y_n^A$, $J \neq I$, $b_n \neq a_n$.



Рис. 1. Базовая схема организации системы машинного анализа массивов графических данных.

Переход от набора матриц $A: \{A_n\}$ к набору матриц $B: \{B_n\}$ является основной частью процесса предварительной обработки и настройки параметров входных графических данных. Данный переход осуществляется через набор функций $\{F_r\}$, где $r \in [1; R]$, которые базируются на математических моделях шумов и аберраций (в данном исследовании все этапы восстановления исходного изображения обобщаются через суммарную функцию F_r^Z , где $F_r^Z(A_n) = B_n$ для $\forall n \in [1; N]$). Далее цветовая схема и динамический диапазон не изменяются и верхний индекс используется для обозначения фрагмента изображения.

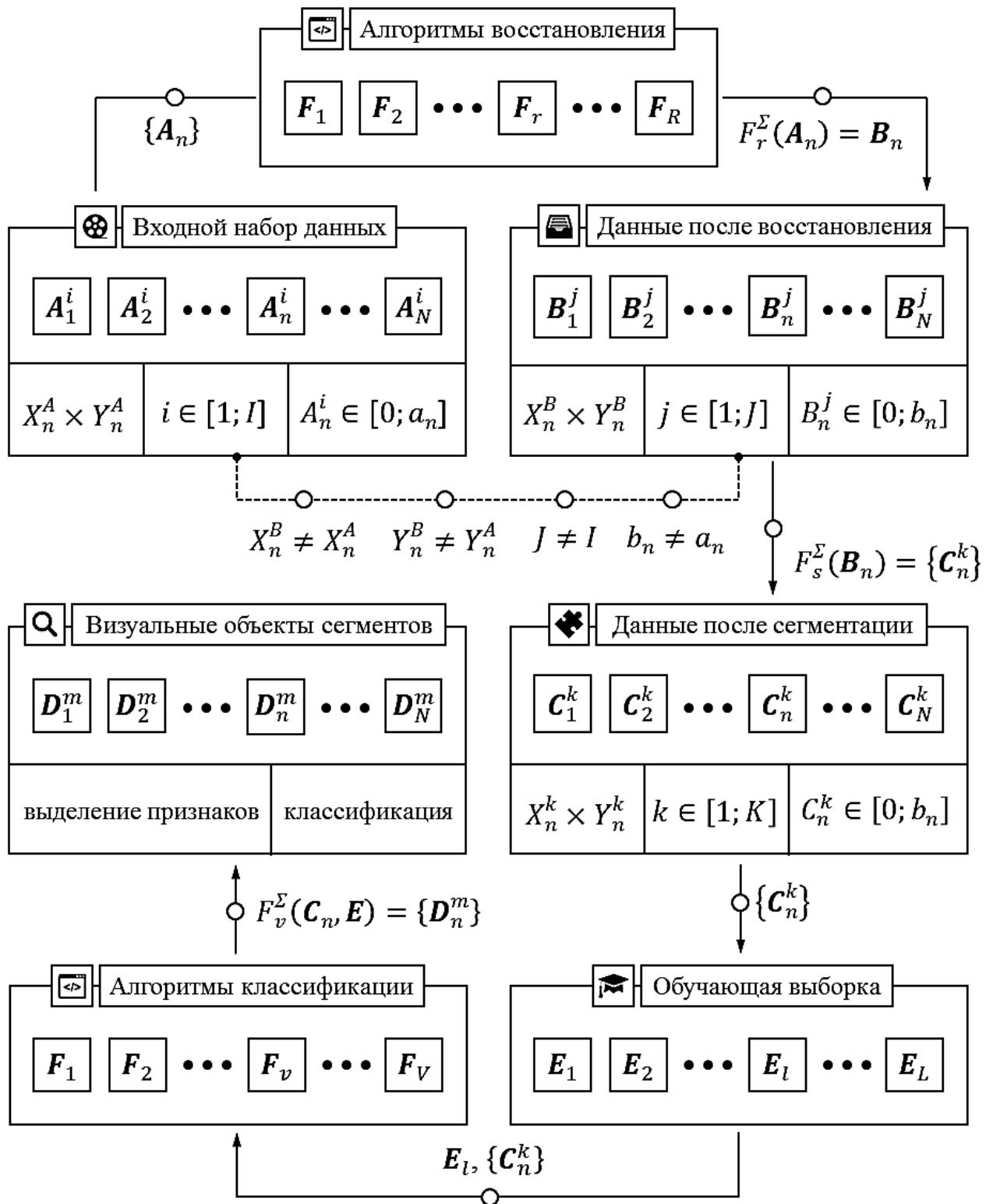


Рис. 2. Математическая модель системы машинного анализа графических данных

Также на этапе предварительной обработки данных проводится сегментация изображения. На уровне построения математического аппарата это формализуется как переход от набора матриц B_n цветных каналов отдельного изображения к набору матриц сегментов $C_n: \{C_n^k\}$, где $k \in [1; K_n]$, как матриц размерности $X_n^k \times Y_n^k$, с прежним динамическим диапазоном, что выражается как $C_n^k \in [0; b_n]$ для $\forall k$ и $\forall n$. Т.е. количество сегментов, как и разрешающая способность каждого сегмента, могут быть различными в каждом отдельном изображении. Как и на предыдущем этапе, сегментация формализуется через набор функций $\{F_s\}$, где $s \in [1; S]$, обобщаемый на основе суммарной функции сегментации F_s^Σ , где $F_s^\Sigma(B_n) = \{C_n^k\}$.

На основе набора восстановленных и скорректированных по разрешению и цветовой схеме изображений $\{B_n\}$, которые были предварительно сегментированы наборы матриц $\{\{C_1^k\}, \{C_2^k\}, \dots, \{C_n^k\}, \dots, \{C_N^k\}\}$ могут быть выделены визуальные объекты $D_n: \{D_n^m\}$, где $m \in [1; M_n]$ у которых выделяются признаки и проводится классификация в соответствии с классификационной схемой или обучающим набором, представленным через

множество $E: \{E_l\}$, где $l \in [1; L]$. В рамках данного исследования, этап классификации также формализуется через набор функций $\{F_v\}$, где $v \in [1; V]$, обобщаемый на основе суммарной функции F_v^E , где $F_v^E(C_n, E) = \{D_n^m\}$.

Таким образом, задача оптимизации системы машинного анализа может быть сведены к математической задаче поиска экстремумов целевых функций (target functions, TF) эффективности классификации. Набор TF определяется на этапе постановки задачи, и в общем случае включают в себя поиск:

- максимума уровня точности выделения визуальных объектов $\max(F_{O}^T)$, выделения признаков визуального объекта $\max(F_{OA}^T)$ и классификации визуального объекта $\max(F_{OC}^T)$;

- минимума отношения времени обработки графических данных к объему данных $\min\left(\frac{\Delta T_A}{V_T}\right)$.

Аргументами FT в данном случае выступают ключевые параметры аппаратно-программного комплекса (вычислительный ресурс и пропускающая способность информационных каналов инфраструктуры системы машинного анализа), а также параметры настройки алгоритмов обработки данных (наборы функций $\{F_r\}$ для $r \in [1; R]$, $\{F_s\}$ для $s \in [1; S]$ и $\{F_v\}$ для $v \in [1; V]$, а также обучающий набор $E: \{E_l\}$ для $l \in [1; L]$).

Таким образом, на основе математических алгоритмов поиска глобальных экстремумов через варьирование указанных параметров вычисляются максимумы и минимумы FT, на основе чего проводится оптимизация системы в соответствии с поставленной задачей.

Выводы

В результате проведенного исследования были проанализированы современные подходы, применяемых при построении алгоритмов выделения и классификации визуальных объектов в массивах графических данных. При этом была указана важность задачи построения комплексной методики для организации систем машинного анализа изображений. В предложенной схеме рассматриваются алгоритмы предварительной обработки входных данных через восстановление исходного изображения, настройку параметров и сегментации, а также алгоритмы выделения визуального объекта, определения его признаков и дальнейшей классификации. Методика оптимизации алгоритмов машинного анализа при этом базируется на математическом моделировании системы регистрации и алгоритмов обработки изображения. В результате, путем построения целевых функций через формализацию задачи по работе с графическими данными проблема оптимизации может быть решена на базе вычисления глобальных экстремумов для зависимостей показателей эффективности выделения и классификации визуальных объектов.

Список литературы / References

1. Zou J., Li Z. & Hong D. (2019). Super-Resolution Reconstruction of Images Based on Microarray Camera. *Computers, Materials & Continua*, 60 (1), 163-177. doi: 10.32604/cmc.2019.05795.
2. Rowcliffe J.M., Carbone C., Jansen P.A., Kays R., Kranstauber B. (2011). Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*. 2 (5). 464–476. doi: 10.1111/j.2041-210x.2011.00094.
3. Bonneau J. (2020). Financial cryptography and data security: 24th international conference, Fc 2020, Kota Kinabalu, Malaysia, February 10-14, 2020: revised selected papers. Springer.
4. Bracciali A., Clark J., Pintore F., Ronne P.B. & Sala M. (2020). Financial Cryptography and Data Security Fc 2019 International Workshops, Voting and Wtsc, St. Kitts, St. Kitts and Nevis, February 18-22, 2019, Revised Selected Papers. Springer International Publishing.
5. Metheny M. (2013). Security and Privacy in Public Cloud Computing. *Federal Cloud Computing*, 71-102. doi:10.1016/b978-1-59-749737-4.00004-6.
6. Cheng P. & Qu H. (2014). Design and Realization Based on Cloud Stack Hybrid Cloud Computing Platform. *Advanced Materials Research*, 989-994, 2297-2300.
7. Seelamantula C.S. & Blu T. (2015). Image denoising in multiplicative noise. 2015 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2015.7351056.
8. Hou J., Tian J. & Liu J. (2005). Spatial image filtering based on wavelet thresholding denoising. *MIPPR 2005: Image Analysis Techniques*. doi:10.1117/12.652332.
9. Zhang C., Liu W. & Xing W. (2018). Color image enhancement based on local spatial homomorphic filtering and gradient domain variance guided image filtering. *Journal of Electronic Imaging*, 27(06), 1. doi:10.1117/1.jei.27.6.063026.
10. Huang C. & Nguyen M. (2019). X-Ray Enhancement Based on Component Attenuation, Contrast Adjustment, and Image Fusion. *IEEE Transactions on Image Processing*, 28(1), 127-141. doi:10.1109/tip.2018.2865637.
11. Petro A. & Sbert C. (2013). Selective Contrast Adjustment by Poisson Equation. *Image Processing On Line*, 3, 208-222. doi:10.5201/ipol.2013.41.
12. Karim T. & Tasneem T. (2014). Analytical Adjustment of Image Contrast. *International Journal of Computer Applications*, 98(20), 44-49. doi:10.5120/17303-7794.
13. Bajla I., Soukup D. & Stolic S. (2011). Occluded Image Object Recognition using Localized Nonnegative Matrix Factorization Methods. *Object Recognition*. doi:10.5772/14124.
14. Han S. & Vasconcelos N. (2006). Image Compression using Object-Based Regions of Interest. 2006 International Conference on Image Processing. doi:10.1109/icip.2006.313095.
15. Adam B., Zaman F., Yassin I., Abidin H. & Rizman Z. (2018). Performance evaluation of faster R-CNN on GPU for object detection. *Journal of Fundamental and Applied Sciences*, 9 (3S), 909.
16. Shiga M., Muto S. (2019). Non-negative Matrix Factorization and Its Extensions for Spectral Image Data Analysis. *E-Journal of Surface Science and Nanotechnology*. 17 (1). 148-154. doi:10.1380/ejssnt.2019.148.

17. Azawi N., Gauch J. (2019). Ransac Based Motion Compensated Restoration for Colonoscopy Images. *Signal & Image Processing: An International Journal*. 10 (4). 9-16. doi:10.5121/sipij.2019.10402.
18. Viola P. and Jones M.J. «Robust real-time face detection», *International Journal of Computer Vision*. Vol. 57, № 2, 2004. Pp. 137–154.
19. Wang X., Ma H. & Chen X. (2016). Salient object detection via fast R-CNN and low-level cues. 2016 IEEE International Conference on Image Processing (ICIP).
20. Alorf A.A. (2016). Performance evaluation of the PCA versus improved PCA (IPCA) in image compression, and in face detection and recognition. 2016 Future Technologies Conference (FTC). doi:10.1109/ftc.2016.7821659.
21. Gnouma M., Ladjailia A., Ejbali R., Zaied M. (2018). Stacked sparse autoencoder and history of binary motion image for human activity recognition. *Multimedia Tools and Applications*, 78 (2), 2157-2179. doi:10.1007/s11042-018-6273-1.
22. Qi X., Silvestrov S., Nazir T. (2017). Data classification with support vector machine and generalized support vector machine. doi:10.1063/1.4972718.
23. Ren Y., Tang L. (2019). A nonconvex and nonsmooth anisotropic total variation model for image noise and blur removal. *Multimedia Tools and Applications*, 79 (1-2), 1445-1473. doi:10.1007/s11042-019-08179-8.
24. Vincent L. & Heijmans H. (2018). Graph Morphology in Image Analysis. *Mathematical Morphology in Image Processing*, 170-203. doi:10.1201/9781482277234-6.
25. Roerdink J.B. (2018). Mathematical Morphology with Noncommutative Symmetry Groups. *Mathematical Morphology in Image Processing*, 205-254. doi:10.1201/9781482277234-7.
26. Liu K., Lu B., Wei Y. (2013). Better image texture recognition based on SVM classification. *MIPPR 2013: Pattern Recognition and Computer Vision*. doi:10.1117/12.2031539.
27. Budiman A., Fanany M.I. & Basaruddin C. (2014). Stacked Denoising Autoencoder for feature representation learning in pose-based action recognition. 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE). doi:10.1109/gcce.2014.7031302.
28. Liang J., & Liu R. (2015). Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. 2015 8th International Congress on Image and Signal Processing (CISP). doi:10.1109/cisp.2015.7407967.
29. Menezes J., Poojary N. (2019). Hyperspectral image Data Classification with Refined Spectral-Spatial features based on Stacked Autoencoder approach. *Recent Patents on Engineering*, 13. doi: 10.2174/187221211366619091114 1616.
30. Sercu T. & Goel V. (2016). Advances in Very Deep Convolutional Neural Networks for LVCSR. *Interspeech 2016*. doi:10.21437/interspeech.2016-1033.
31. Zhang Z., Gong C., Liu R. (2017). Face Detection Based on Method Combined RVM and SVM. *Computer Science and Artificial Intelligence*. doi:10.1142/9789813220294_0058.
32. Ganakwar D.G., Kadam V.K. (2019). Face Detection Using Boosted Cascade of Simple Feature. 2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC). doi:10.1109/icraecc43874.2019.8994977.