ThangPQ, LamHT 2023, 11:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Deep Learning and RBF Hybrid Models for Flower Image Recognition

Pham Quoc Thang, Hoang Thi Lam
Tay Bac University, Vietnam

Abstract-Image object recognition is easy for humans, but a complicated problem for machines. The purpose of flower image recognition is to determine the suitable flower species for the input image, based on the features. In recent years, deep learning (DL) models have been widely and successfully applied in many fields. In this paper, we propose and study the feasibility and effectiveness of general CNN-RBF hybrid models for flower image recognition problem. The experimental results on two flower image datasets, Oxford-17 and Oxford-102 flowers, show that the CNN-RBF hybrid models in general, especially the CNN-SVM hybrid model, give better recognition results than original CNN model and can be applied to effectively classify flower images.

Keywords-Deep Learning, RBF, Hybrid Model, Image Recognition.

I. INTRODUCTION

Image object recognition is easy for humans, but a complicated problem for machines. The recognition system consists of the following main components: image sensor (or camera, to record images), image preprocessing, object detection, object segmentation, feature extraction, and object classification. The recognition system consists of a classified image sample database, which is used for training. When an image object appears, it will be classified into an appropriate class.The purpose of flower image recognition is to determine the suitable flower species for the input image, based on the features. The first task in flower image recognition is to extract image features and the second task is to classify flower images into a suitable class. To extract features of flower image, there are several methods such as Fourier transform [40], Wavelet [23], Hough [34], curvelet and ridgelet [33] or principal component analysis (PCA) [12], independent component analysis. (ICA) [39]... In [18], the authors use DeepCNN model to extract features of flower image. Each method has its own strengths and weaknesses. The user needs to choose the appropriate method for his problem of interest.There are many techniques for classifying flower images such as k-Nearest Neighbor (K-NN) [17], Bayes Network [25], Adaptive boost (Adaboost) [11], Artificial Neural Network (NN) [25] and Support Vector Machine (SVM) [25].

In this paper, we propose a general CNN-RBF hybrid model using CNN to extract features of flower images, then use efficient RBF models (SVM, RVM) for classification. We experimentally classify flower images to show that the CNN-RBF hybrid models in general, especially the CNN-SVM model, give better recognition results than the original CNN model.
The rest of this paper is structured as follows: Part II presents RBF models such as SVM, RVM. Part III presents the hybrid models of deep learning and

RBF. Part IV describes the flower image recognition problem along with the experimental results and conclusions are presented in the last part.

II. RBF MODELS

1. Introduction to The RBF Model

Radial basis function (RBF) model is a basic model that has been used in solving many different problems and is continuing to be widely applied in many different practical applications such as classification, function approximation, predict data over time [2] [26]... RBF model has a simple structure, in the form of a linear combination of basis functions:

$$f(x;w) = \sum_{i=1}^{M} w_i \phi_i(x) + b \qquad (1)$$

Here, it is common to use the radial basis function (RBF) $\phi_i(x)$ = $\exp(-\gamma \|x - c_i\|^2)$, i = 1,.. , M depends only on the distance from the argument x to a given point ci (called the center) with width ･ and M is the number of radial basis functions ･i of the model is used to calculate the function f. In fact, the RBF model achieves good classification performance in the applications of image recognition [15], speech recognition [1] and human gesture recognition [14]...

To find a decision function of the form (1), there are many approaches expressed through different objective functions, in each approach there are many different solutions. The following sections will present two approaches to building and using RBF models in classification problems, namely SVM and RVM.

2. Support Vector Machines

SVM (Support Vector Machines) works in feature space F via a kernel function $K(x, y) = \Phi(x).\Phi(y)$ where $\Phi : R^d \rightarrow F$ is a map from the d-dimensional input space to a possibly high-dimensional feature space [3]. For a two-class classification problem, the decision rule takes the form:

$$y = sign\left(\sum_{i=1}^{M} \alpha_i K(x, x_i) + b \right) \qquad (2)$$

where α_i are weights of support vectors x_i, x is the input vector needed to classify,

$$K(x, y) = \Phi(x).\Phi(\mathrm{y}) \text{, } K(x, x_i) = \Phi(x).\Phi(\mathrm{x}_i) \qquad (3)$$

is a kernel function calculating the dot product of two vectors $\Phi(x)$ and $\Phi(x_i)$ in the feature space, b is the bias, and M is the number of support vectors. The task of the SVMs training process is to determine all the parameters (x_i, α_i, b, M); the resulting x_i, $i = 1,..,M$ are a subset of the training set and are called support vectors.

3. Relevance Vector Machines

Another approach in constructing the decision function (1) is based on the Bayesian inference principle. Given a two-class dataset $T = \{(x_i,y_i), x_i \in R^d , yi \in \{0,1\}, i = 1,...,n\}$, the RVM (Relevance Vector Machines) method [30] uses an assumption that y has a Bernoulli distribution and y_i are independent. Likelihood of the training dataset for the parameters w_i is:

$$P(y \mid w) = \prod_{i=1}^{n} \delta(f(x_i))^{y_i} (1 - \delta(f(x_i)))^{1-y_i} \qquad (4)$$

in which $\delta(y) = 1/(1 + e^{-y})$ is a logistic function whose input is the value of the linear function (1).

To limit the number of components $w_i \neq 0$, Mike Tipping [30] uses additional prior constraints that each parameter has a normal distribution with a mean of 0 and a hyperparameter α_i for the variance:

$$p(w \mid \alpha) = \prod_{i=1}^{n} N(w_i \mid 0, \alpha_i^{-1}) \qquad (5)$$

Constraint (5) is set up for two purposes. The first purpose is to make model (1) simpler to avoid over-fitting in training. The second purpose is that model (1) will run faster by using fewer basis functions.

III. Deep Learning and RBF Hybrid Models

In recent years, deep learning (DL) models have become the mainstream in big data analysis, have been widely and successfully applied in many fields such as: image recognition [8], speech recognition [10], natural language processing [22], disease diagnosis [27]... due to its superior performance compared to traditional machine learning models.

1. Deep Learning Model in Classification

In previous machine learning models, manually extracted features, sometimes called "shallow features", were extracted based on domain-specific knowledge. Extracting them is time consuming and often difficult to apply to some types of data such as raw images... Deep learning models are capable of automatically extracting features from the input data, for example such as raw images... These features are considered abstract and high-level features, which are often more efficient for classification than shallow features. The high-level learning features extracted in deep learning networks have proved very effective in machine vision, speech processing...

Different image classification techniques at the output layer in deep learning networks with different structure. Yuan [36] uses a deep belief network (DBN) multilayer structure to learn visual features and tag features in images. Tang [29] solves histogram classification by using DBN for features extracted by CNN network. The features extracted in the deep learning model are quite effective when dealing with large data sets. In [9], Krizhevsky trained a large CNN network to classify 1.2 million high-resolution images into 1000 classes in ImageNet LSVRC. Image classification applications that use CNN for feature extraction have high performance. The deep features extracted in the CNN network outperform the manually extracted shallow features (Figure 1).

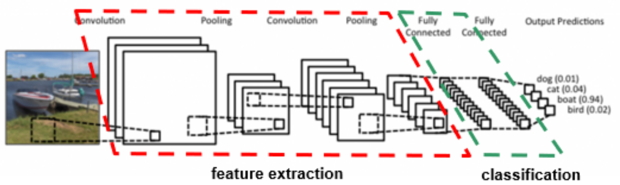

Figure 1: The CNN model is used to classify images

Niu [20] proposes a multi-layer CNN with a combined deep architecture of automatic feature extractor and classifier. The automatic feature extractor has mapping and feature extraction layers from the image corresponding to two operations: convolutional filters and pooling. The advantage of the CNN classifier is the automatic extraction of the features of the input image. These features are usually invariant to the geometrical displacements and distortions of the input data. Meanwhile, "shallow" feature extraction is laborious, having to apply many different types of features to obtain invariance to the geometric deformation of the input data.

Typically, image classification applications use the features of the last layer in deep learning networks for classification. The last layer is quite sensitive to semantic information, while the middle layers are less sensitive to semantics, but have the ability to preserve more detail. The layers represent the hierarchy of features [37]. Therefore, it is possible to view different CNN layers corresponding to different levels of abstraction.

2. CNN-SVM Hybrid Model

Niu [20] proposes a CNN-SVM hybrid model in handwriting recognition (Figure 2). The CNN-SVM hybrid model combines the synergy of two CNN and SVM classification models. The architecture of the CNN-SVM hybrid model is designed by replacing the final output layer of the CNN with an SVM classifier. In this model, the CNN acts as the feature extractor, and the SVM acts as the classifier. Hybrid model allows automatically extracting features from raw images, then classifying them using SVM.

Omaraa [21] offers a multimedia biometric classification and recognition system for face and ear images. The author proposes a way to exploit features extracted from CNN on face and ear images, giving features that allow strong discrimination. First, the features of the face and ear images are extracted based on the VGG-M network. Next, the features are merged using DCA technique and finally, classified using SVM.

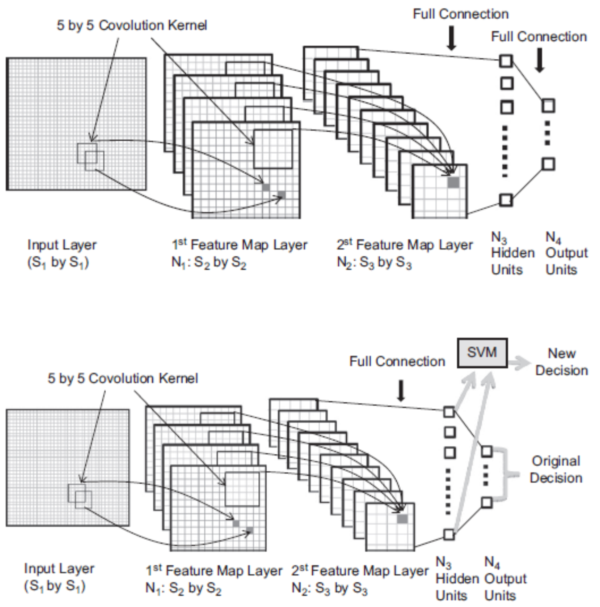

Figure 2: CNN-SVM hybrid model for handwriting recognition

3. Proposing a CNN-RBF General Hybrid Model

The paper proposes a CNN-RBF general hybrid model in which the CNN-SVM hybrid model [20] with the architecture shown in Figure 2 is a specific case. The CNN-RBF general hybrid model (Figure 3) replaces the final output layer of the CNN with a classifier based on the RBF model such as SVM, RVM... Then, the output values of the last hidden layer will only make sense to CNN itself. However, these values can be considered as features for SVM, RVM classifiers... In other words, in the CNN-RBF general hybrid model proposed by the paper, the output values at the last layer of CNN are input features for classifier based on RBF model such as SVM, RVM... In this model, CNN acts as feature extractor and RBF model such as SVM, RVM... acts as a classifier.

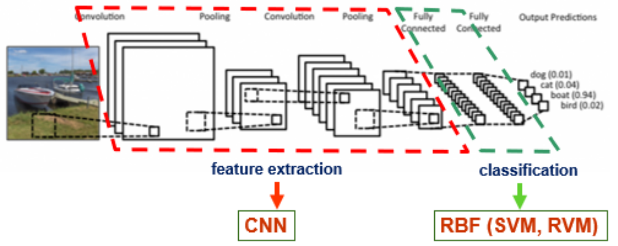

Figure 3: CNN-RBF general hybrid model

The proposed CNN-RBF general hybrid model is a promising classification model because it has two characteristics:

+ The features are automatically extracted by CNN, while most of the traditional classification models are based on manual feature extraction which is quite laborious and time consuming.

+ The CNN-RBF general hybrid model (CNN-SVM, CNN-RVM...) combines the advantages of CNN and SVM, RVM, which are popular and effective classification models.

IV. EXPERIMENT

1. Data Set

The paper uses two image datasets of common flowers in the United Kingdom, Oxford 17-Flowers and Oxford 102-Flowers of the University of Oxford [19]. The Oxford 17-Flowers dataset includes 1360 images of 17 species of flowers, averaging 80 images for each flower. The Oxford 102-Flowers dataset includes 7370 images of 102 flower species, each with between 40 and 200 images. The images on both datasets varied in size, angle, and light variation. Image classes have quite strong variation of images in the same class or images that are close to images in other classes. Some examples of flower images are shown in Figure 4.

Figure 4: Illustrate examples of flower images in Oxford 17-Flowers and Oxford 102-Flowers datasets

2. Feature Extraction

In flower image classification, the data are images of flowers. Therefore, instead of using shallow feature extraction methods in traditional machine learning models, to classify flower images, the paper uses CNN to automatically extract flower image features. The paper using the ResNet18 network is a pre-trained CNN using a large dataset of ImageNet rich images with about 1000 image classes and 1.2 million training images [5]. From this large ImageNet data set, CNN can learn rich features representing many types of images. These features often represent better than manually extracted features such as HOG, LBP or SURF.

To extract features of two Oxford 17-Flowers and Oxford 102-Flowers datasets, the paper uses ResNet18 as a CNN to extract features. The paper refines the ResNet18 network structure, keeping only the ConvNet layers in the CNN and omitting the FC layers (Figure 5) and uses the output of the remaining ConvNet layers as input features to classify using the RBF models such as SVM, RVM. Then, re-train the refined ResNet18 network with the image data of two Oxford 17-Flowers and Oxford 102-Flowers data sets. The result is a retrained CNN used to extract features from flower image data.

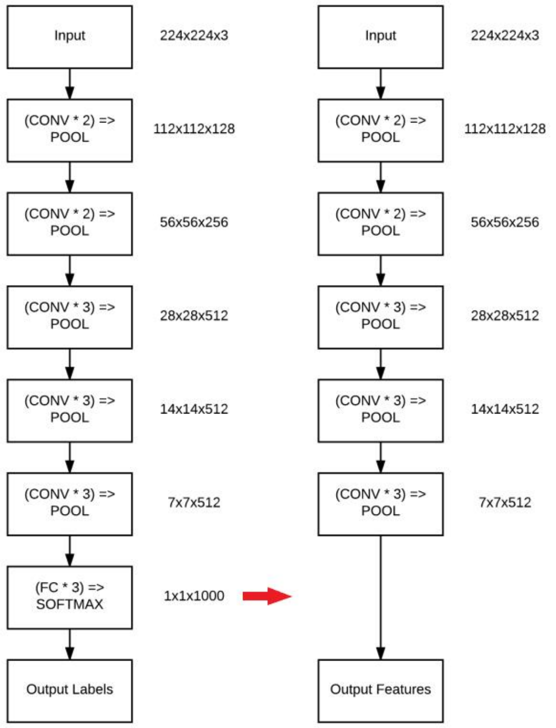

Figure 5: The model uses CNN for automatic feature extraction of flower images

3. Flower Image Classification

The specific experimental results are given in Table 1. From those experimental results, the following points can be observed:

+ About recognition accuracy: the CNN-RBF hybrid models in general, especially the CNN-SVM model, give better recognition results compared to the original CNN model. This can be explained because CNN-SVM hybrid models combine the advantages of features automatically extracted by CNN deep learning network with the power of SVM classifier.

The CNN-RVM hybrid models have lower recognition accuracy than the CNN-SVM hybrid models on the feature set learned through the ResNet deep learning network of two Oxford 17-Flowers and Oxford 102-Flowers datasets in the flower image classification task. This is known, because the CNN-RVM model is more compact, with fewer basis vectors, so it can reduce the prediction accuracy more. However, with a smaller number of basis vectors in the model, the CNN-RVM hybrid models will achieve faster prediction speed in the test phase.

+ About the training time: the training time is also measured to compare the speed of the algorithms under consideration. It can be seen that the CNN-SVM hybrid model has a training time that can be tens of times faster than the CNN-RVM hybrid model, but still ensures accuracy.

Table -1: Flower image classification results.

Methods	Oxford 17-Flowers			Oxford 102-Flowers		
	Acc(%)	SVs	Time(s)	Acc(%)	SVs	Time(s)
CNN	96.69	-	-	96.95	-	-
CNN-SVM	97.43	659	0.61	97.22	4589	28.69
CNN-RVM	96.69	65	46.8	95.59	509	11449

+ The paper also compares the above experimental results with recent research results of other authors on the same two Oxford 17-Flowers and Oxford 102-Flowers data sets in Table 2. The CNN-SVM hybrid model has higher recognition accuracy results than many other methods, only slightly lower than the results in [18] and [28]. However, in [18], the authors calculate the classification accuracy on the training dataset. While the paper measures the classification accuracy on the test dataset independent of the training data set, so it has a more independent, objective, and difficult difficulty. In [28], the authors use EfficientNet-B7 network for feature extraction, this is a new, more efficient network architecture than the ResNet18 network that the paper uses, so the learned features of [28] can be better. . These results show that CNN-SVM, CNN-RVM hybrid models can be applied to classify flower image recognition effectively.

IV. CONCLUSION

We present our research on the feasibility and effectiveness of the deep learning and RBF hybrid model when applied to flower image recognition problem. Our experimental results show that the CNN-RBF hybrid models in general, especially the CNN-SVM model, give better recognition results than the original CNN model, which is very competitive compared with other methods and can be applied to classify flower image recognition effectively.

Table -2: Compare the results of flower image classification methods

Research work	Classification results (Accuracy %)	
	Oxford 17-Flowers	Oxford 102-Flowers
Xiaoling et al, [35] (Using Inception-v3 network)	95%	94%

Mete et el, [18] (Using GoogleNet in conjunction with SVM)	99.6%	98.5%
Mete et el, [18] (Using GoogleNet in conjunction with MLPC)	99.8%	96.5%
Mete et el, [18] (Using GoogleNet in conjunction with KNN)	99.1%	96.2%
Mete et el, [18] (Using GoogleNet in conjunction with RF)	95.4%	87.6%
Tan et el, [28] (Using EfficientNet-B7 network)	-	98.8%
Dubey et el, [6] (Using PC Bilinear CNN)	-	93.65%
Cubuk et el, [4] (Using AutoAugment)	-	95.36%
Touvron et al, [32] (Using FixInceptionResNet-V2)	-	95.7%
Touvron et el, [31] (Using ResMLP-12)	-	97.4%
Touvron et el, [31] (Using ResMLP-24)	-	97.9%
Lu et al, [16] (Using NAT-M2)	-	97.9%
Lin et all, [13] (Using Deep CNN)	96.84%	90.85%
Shi et al, [24] (Using Inception-V3)	95%	95%
Hosseini et al, [7] (Using MKL)	93.8%	-
Zhang et al, [38] (Using DCNN)	-	97.34%
CNN	96.69%	96.95%
CNN-SVM	97.43%	97.22%
CNN-RVM	96.69%	95.59%

REFERENCES

1. Ayrancı, A. A., Atay, S., Yıldırım, T., "Speaker Accent Recognition Using Machine Learning Algorithms", in: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1–6.
2. Baklacioglu, T. (2021), "Predicting the fuel flow rate of commercial aircraft via multilayer perceptron, radial basis function and ANFIS artificial neural networks", The Aeronautical Journal, 125, pp. 1–19.
3. Cortes, C., Vapnik, V. (1995), "Support-Vector Networks", Mach. Learn., 20 (3), 273–297
4. Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., Le, Q. V., "AutoAugment: Learning Augmentation Strategies From Data", in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 113–123.
5. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., "ImageNet: A large-scale hierarchical image database", in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
6. Dubey, A., Gupta, O., Guo, P., Raskar, R., Farrell, R., Naik, N., "Pairwise Confusion for Fine-Grained Visual Classification", in: Computer Vision – ECCV 2018, Cham: Springer International Publishing, 2018, pp. 71–88.
7. Hosseini, B., Hammer, B. (2019), "Large-Margin Multiple Kernel Learning for Discriminative Features Selection and Representation Learning", 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.
8. Islam, M. T., Karim Siddique, B. N., Rahman, S., Jabid, T., "Image Recognition with Deep Learning", in: 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), vol. 3, 2018, pp. 106–110.
9. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "ImageNet Classification with Deep Convolutional Neural Networks", Commun. ACM, 60 (6), 84–90.
10. Lakkhana wannakun, P., Noyunsan, C., "Speech Recognition using Deep Learning", in: 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 2019, pp. 1–4.
11. Li, S., Zhu, L., Jiang, T., "Active Shape Model Segmentation Using Local Edge Structures and AdaBoost", in: Medical Imaging and Augmented Reality, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 121–128.
12. Lihong, Z., Zikui, G., "Face Recognition Method Based on Adaptively Weighted Block-Two Dimensional Principal Component Analysis", in: 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, 2011, pp. 22–25.
13. Lin, K, Yang, H., Chen, C., "Flower classification with few training examples via recalling visual patterns from deep CNN", in: IPPR Conference on Computer Vision, Graphics, and Image Processing (CVGIP), 2015, 1–8.
14. Liu, F., Zeng, W., Yuan, C., Wang, Q., Wang, Y. (2019), "Kinect-based hand gesture recognition using trajectory information, hand motion dynamics and neural networks", Artificial Intelligence Review, 52, pp. 563–583.
15. Liu, Y. (2019), "Digital Image Recognition Based on Improved Cognitive Neural Network", Translational Neuroscience, 10, pp. 125–128.
16. Lu, Z., Sreekumar, G., Goodman, E., Banzhaf, W., Deb, K., Boddeti, V. (2021), "Neural Architecture Transfer", IEEE Transactions on Pattern Analysis and Machine Intelligence, 43 (09), pp. 2971–2989.
17. McSherry, D., Stretch, C., "An Analysis of Order Dependence in kNN", in: Artificial Intelligence and Cognitive Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 207–218.

18. Mete, B. R., Ensari, T., "Flower Classification with Deep CNN and Machine Learning Algorithms", in: 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019, pp. 1–5.
19. Nilsback, M.-E., Zisserman, A., "Automated Flower Classification over a Large Number of Classes", in: 2008 Sixth Indian Conference on Computer Vision, Graphics Image Processing, 2008, pp. 722–729.
20. Niu, X.-X., Suen, C. Y. (2012), "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", Pattern Recognition, 45 (4), pp. 1318–1325.
21. Omara, I., Xiao, G., Amrani, M., Yan, Z., Zuo, W., "Deep features for efficient multi-biometric recognition with face and ear images ", in: Ninth International Conference on Digital Image Processing (ICDIP 2017), vol. 10420, International Society for Optics and Photonics, SPIE, 2017, pp. 68 –73.
22. Otter, D., Medina, J. R., Kalita, J. (2021), "A Survey of the Usages of Deep Learning for Natural Language Processing", IEEE Transactions on Neural Networks and Learning Systems, 32, pp. 604–624.
23. Shen, L., Bai, L., Picton, P., "Facial recognition/verification using Gabor wavelets and kernel methods", in: 2004 International Conference on Image Processing, 2004. ICIP '04. Vol. 3, 2004, 1433–1436 Vol. 3.
24. Shi, L., Li, Z., Song, D. (2019), "A Flower Auto-Recognition System Based on Deep Learning", IOP Conference Series: Earth and Environmental Science, 234, p. 012088.
25. Shukla, A., Agarwal, A., Pant, H., Mishra, P. (2020), "Flower Classification using Supervised Learning", International Journal of Engineering Research and, V9, pp. 757–762.
26. Shymkovych, V., Telenyk, S., Kravets, P. (2021), "Hardware implementation of radial-basis neural networks with Gaussian activation functions on FPGA", Neural Computing and Applications, 33 (15), pp. 9467–9479.
27. Sidey-Gibbons, J., Sidey-Gibbons, C. (2019), "Machine learning in medicine: a practical introduction", BMC Medical Research Methodology, 19, 19:64.
28. Tan, M., Le, Q. V., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", in: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA, vol. 97, Proceedings of Machine Learning Research, PMLR, 2019, pp. 6105–6114.
29. Tang, B., Liu, X., Lei, J., Song, M., Tao, D., Sun, S., Dong, F. (2016), "DeepChart: Combining deep convolutional networks and deep belief networks in chart classification", Signal Processing, 124, Big Data Meets Multimedia Analytics, pp. 156–161.
30. Tipping, M. E., "The Relevance Vector Machine", in: Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99, Denver, CO: MIT Press, 1999, 652–658.
31. Touvron, H., Bojanowski, P., Caron, M., Cord, M., El-Nouby, A., Grave, E., Joulin, A., Synnaeve, G., Verbeek, J., Jégou, H. (2021), "ResMLP: Feedforward networks for image classification with data-efficient training", CoRR, abs/2105.03404, pp. 1–12.
32. Touvron, H., Vedaldi, A., Douze, M., Jégou, H. (2020), "Fixing the traintest resolution discrepancy: FixEfficientNet", CoRR, abs/2003.08237, pp. 1–3.
33. Wakin, M. B. (2011), "Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity (Starck, J.-L., et al; 2010) [Book Reviews]", IEEE Signal Processing Magazine, 28 (5), pp. 144–146.
34. White, K. P., Kundu, B., Mastrangelo, C. M. (2008), "Classification of Defect Clusters on Semiconductor Wafers Via the Hough Transformation", IEEE Transactions on Semiconductor Manufacturing, 21 (2), pp. 272–278.
35. Xia, X., Xu, C., Nan, B., "Inception-v3 for flower classification", in: 2017 2nd International Conference on Image, Vision and Computing (ICIVC), 2017, pp. 783–787.
36. Yuan, Z., Sang, J., Xu, C., "Tag-aware image classification via Nested Deep Belief nets", in: 2013 IEEE International Conference on Multimedia and Expo (ICME), 2013, pp. 1–6.
37. Zeiler, M. D., Fergus, R., "Visualizing and Understanding Convolutional Networks", in: Computer Vision – ECCV 2014, Cham: Springer International Publishing, 2014, pp. 818–833.
38. Zhang, R., Tian, Y., Zhang, J., Dai, S., Hou, X., Wang, J., Guo, Q. (2021), "Metric learning for image-based flower cultivars identification", Plant Methods, 17, pp. 1–14.
39. Zhang, X., Ren, X., "Two Dimensional Principal Component Analysis based Independent Component Analysis for face recognition", in: 2011 International Conference on Multimedia Technology, 2011, pp. 934–936.
40. Zhang, Y., He, R., Jian, M., "Comparison of Two Methods for Texture Image Classification", in: 2009 Second International Workshop on Computer Science and Engineering, vol. 1, 2009, pp. 65–68.