Intelligent Fusion of Deep Features for Improved Waste Classification

KASHIF AHMAD[ID]1, **KHALIL KHAN**[ID]2, **AND ALA AL-FUQAHA**[ID]1, **(Senior Member, IEEE)**
1Information and Computing Technologies (ICT) Division, College of Science and Engineering (CSE), Hamad Bin Khalifa University, Doha, Qatar
2Department of Electrical Engineering, The University of Azad Jammu & Kashmir, Muzaffarabad 13100, Pakistan

Corresponding author: Kashif Ahmad (kahmad@hbku.edu.qa)

ABSTRACT In this article, we address the problem of an image-based automatic classification of waste materials. Given the large number of waste categories and the importance of proper management of waste materials, the problem is known to be critical and of a particular interest. To achieve reliable waste classification capability, we propose a novel approach, that we name double fusion, which optimally combines multiple deep learning models using feature and score-level fusion methods. The double fusion scheme ensures an optimized contribution of the deep models by, firstly, combining their capabilities in an early and late fusion scheme followed by a score-level fusion of the classification results obtained with early and late fusion methods. In total, we employ and compare six different fusion methods including two feature-level fusion schemes, namely (i) Discriminant Correlation Analysis and (ii) simple concatenation of deep features, and four late fusion methods, namely (i) Particle Swarm Optimization, (ii) Genetic modeling of deep features (iii) Induced Ordered Weighted Averaging and (iv) a baseline method where all the deep models are treated equally. Moreover, we also evaluate the performance of the individual deep models, and compare our results against state-of-the-art methods demonstrating a significant improvement of 3.58% over state-of-the-art.

INDEX TERMS Waste management, waste classification, deep features, fusion, double fusion, particle swarm optimization, genetic algorithms, IOWA.

I. INTRODUCTION

Waste disposal has a direct or indirect impact on human lives and the environment. A proper waste management system can help in mitigating the adverse effects of waste materials. Waste management involves several activities, such as waste collection, separation/classification, and disposal or recycling. Classification of waste into different categories based on the nature of the materials is one of the key activities of waste management, which may affect the rest of the process [1].

Being a key component of waste management, waste separation and classification has been an area of keen interest for the researchers over the last few years. During this time, several interesting solutions, targeting different aspects of waste classification, have been proposed [2]. For instance, Sander *et al.* [3] review and analyze the European Commission's (EC) list of waste classification,[1] and make recommendations for handling waste in light of EC's guidelines. More recently, a growing interest in Computer Vision (CV) and Machine Learning (ML) communities has been noticed in the task. To this aim, several interesting solutions have been proposed [4], [5]. The majority of the solutions rely on existing deep architectures, mostly fine-tuned or retrained from scratch. For instance, in [6] and [7], Fast R-CNN [8] and GoogleNet [9] have been fine-tuned on garbage images, respectively. However, no ultimate solution is still available. Several aspects need to be investigated yet. For instance, how different models respond to the task and how effectively the responses/features extracted by different models could be combined leading to an improvement in the overall performance.

In this paper, we propose to fuse multiple deep models, exploring the capabilities of several early and late fusion techniques both individually and jointly combined in a novel

The associate editor coordinating the review of this manuscript and approving it for publication was Navanietha Krishnaraj Krishnaraj Rathinam.

[1]https://ec.europa.eu/environment/waste/framework/list.html

double fusion scheme where the contributions of early and late fusion methods are combined in an optimal way for image-based waste classification task. In details, the capabilities of the deep models are firstly combined in several ways including two early fusion strategies, namely Discriminant Correlation Analysis (DCA) and simple concatenation of deep features, and four late fusion methods, namely Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Induced Ordered Weighted Averaging (IOWA) based Fusion and simple averaging of classification scores obtained with classifiers trained on features extracted via different deep models. The classification scores obtained as results of the best individual early and late fusion methods are further combined in a method termed as double fusion using a late fusion method by assigning them merit-based weights optimized via PSO and GA. The motivation for the fusion-based framework is based on the assumption that the existing state-of-the-art deep architectures respond differently to an image by extracting diverse but complementary image features. Combining such diverse but complementary features could result in an improvement in the performance of a waste classification framework. We also assess the performances of the individual deep models employed as features descriptors without any fine-tuning and re-training. To the best of our knowledge, no prior work provides such rigorous analysis by employing multiple existing deep models as feature descriptors for waste classification in the double fusion manner. The motivation for deep features (extracted through existing pre-trained models) based approach is based on its proven performances in different application domains [10]–[12]. Moreover, building and training a custom deep model from scratch requires heavy computation resources as well as large amounts of data. Particularly, in the case of fusion, training or fine-tuning would require more computational resources especially fusing several deep architectures in a single framework. We believe this work will provide a baseline for future research in the domain.

The main contributions of the work can be summarized as:

(i) Being an important component of computer vision and multimedia frameworks, we conduct an analysis and evaluation of several deep features in waste classification tasks.

(ii) We demonstrate that different deep architectures pose diverse and complementary characterization capabilities, and their fusion could result in an improvement in the performance of waste classification frameworks.

(iii) With several fusion strategies including early and late, naive, and merit-based linear and non-linear fusion schemes, we demonstrate that the merit and capabilities of each model should be considered in the fusion process by attributing merit-based weights instead of treating them equally.

(iv) We propose a novel double fusion scheme to combine the capabilities of early and late fusion methods optimally.

(v) We demonstrate significant improvement for the proposed methods over state-of-the-art techniques.

The rest of the paper is organized as follows: Section II provides a detailed overview of the related work. Section III describes the proposed methodology while the details of the dataset, experimental setup, conducted experiments, and achieved results are provided in Section IV. Finally, Section V concludes our work.

II. RELATED WORK

In recent years, image-based waste classification systems have got great attention from the research community. Several interesting solutions employing a diversified set of techniques and technologies have been proposed over the last few years. For instance, Donovan [13] proposed Auto Trash, a Raspberry Pi-powered system equipped with a camera, to automatically sort garbage items based on the nature of the materials. The system relies on a custom software model based on Convolutional Neural Networks (CNNs) built on top of the Google's TensorFlow[2] AI engine. The system in current implementation can only differentiate in compostable and recyclable items. However, it could be extended to other categories. On the other hand, Mittal *et al.* [14] proposed and developed a smartphone application namely SpotGarbage for garbage classification. For classification of the garbage images, the system relies on an existing deep architecture namely AlexNet.

Similarly, deep learning-based approaches have also been used in [7], [15]. In [7], different existing pre-trained models are fine-tuned for the garbage classification. In total, four different state-of-the-art deep models, pre-trained on ImageNet, have been fine-tuned and evaluated on six different categories of garbage. Moreover, a comparative analysis of different classification algorithms including SVMs and Soft-Max used on top of the CNN model has also been conducted. On the other hand, Yang *et al.* [15] rely on a fast R-CNN based framework. Also, the authors evaluated handcrafted features, namely Scale-Invariant Features Transforms (SIFT) and bag of features model for SVMs based classification of garbage items. Moreover, a new garbage dataset containing a total of 2,527 images has also been collected. Another deep architecture based solution has been proposed in [16], where ResNet and DenseNet [17] are retrained from scratch as well as fine-tuned. In the fine-tuning experiments, the existing models pre-trained on ImageNet have been utilized. Moreover, the authors also proposed a novel model namely RecycleNet. Sousa *et al.* [18] proposed a two-step approach for garbage classification in an industrial environment. The solution takes advantage of the Faster R-CNN object detector to detect the objects of interest in images. The evaluation of the method is carried out on a newly collected dataset. Wang *et al.* [19] also collected a new dataset containing plastic and non-plastic garbage related images for the evaluation of their deep architecture based garbage classification system.

---

[2]https://www.tensorflow.org/

There are also some efforts relying on hand-crafted features. For instance, Liu *et al.* [20], relies on a set of low and mid-level features including color, texture, micro-texture, and outline-shape based features. In total, eight different types of features, namely color, SIFT, jet, micro-SIFT, micro-jet, curvature, edge-slice, and edge ribbon, have been employed both individually and in combination using an augmented Latent Dirichlet Allocation (aLDA) model.

The literature depicts that the majority of the proposed solutions rely on deep architecture mostly fine-tuned or retrained from scratch. However, to the best of our knowledge no prior work provides such rigorous analysis by employing multiple existing deep models as feature descriptors for waste classification in different ways including early vs late, naive vs merit-based fusion methods.

III. METHODOLOGY

Figure 1 provides the block diagram of the proposed double fusion framework. The proposed method mainly consists of three phases namely (i) feature extraction, (ii) classification, and (iii) fusion. In the first phase deep features have been extracted via several deep models followed by Support Vectors Machines (SVMs) based classification. Several fusion techniques containing both early and late fusion schemes are then used to jointly utilize these models. Finally, the capabilities of early and late fusion methods are then combined in a double fusion method. The first two phases are rather standard, however, the main strength of the methodology lies in the fusion phase. In the next subsections, we provide a detailed description of the extraction, classification, and fusion schemes used in this work.

A. FEATURE EXTRACTION AND CLASSIFICATION

As mentioned earlier, the work mainly focuses on the fusion phase and the deep models are simply used as feature descriptors. The motivation for using pre-trained CNN models as feature descriptors comes from the fact that building a custom deep model from scratch requires heavy computation resources as well as large amounts of data. Particularly, in the case of fusion, training would require more computational resources especially fusing several deep architectures in a single framework. The literature also depicts good performance for existing deep models when used as feature descriptors in different application domains [10], [11]. We mainly used four state-of-the-art deep architectures, namely (i) AlexNet [21], (ii) GoogleNet [9], (iii) VggNet [22] and (iv) ResNet [23], pre-trained on ImageNet [24]. These models are selected based on their performances in various application domains [11]. AlexNet is composed of 8 weighted layers, GoogleNet consists of 22 layers while VggNet and ResNet are available in multiple configurations, however in this work we opted for the configurations with 19 and 101 layers for VggNet and ResNet, respectively. Features are extracted from the last fully connected layers without any fine-tuning and re-training using Caffe toolbox.[3] AlexNet and VggNet produce feature vectors of 4096 dimensions while ResNet and GoogleNet resulted in 1000 dimensional feature vectors. Table 1 provides a summary of the deep models/architectures used for feature extraction in this work in terms of the number of layers, the feature extraction layer, and the length of the feature vectors obtained with each model.

After extracting features with the pre-trained models, SVMs are trained on features extracted with each deep model. For the implementation of SVMs, we used Fit multiclass model from Mathworks[4] with default parameters (i.e., SVM binary learners).

B. FUSION SCHEMES

For the joint use of the deep models for waste classification, we rely on several fusion schemes including both: early and late fusion methods. In total, we experimented with six fusion schemes. Two of the methods are based on early fusion (i.e., feature level) while four of our fusion techniques are based on late fusion (i.e., classification level). We also combined the classification scores obtained with early and late fusion in a double fusion method where the scores obtained with each fusion scheme are simply combined by assigning both: equal weights and merit-based weights. All the fusion methods are implemented in Matlab. A detailed description of the fusion methods used in this work has been provided in the next sub-sections.

1) EARLY FUSION SCHEMES

For early fusion, we rely on two different techniques including (i) a simple concatenation of feature vectors obtained with the deep models as a baseline method and (ii) DCA for feature-level fusion of the extracted features. In the first method, feature vectors obtained through the individual deep models are simply concatenated, and an SVMs classifier has been trained on the resultant feature vectors. Our second early fusion method is based on DCA based feature-level fusion by exploiting class associations in the correlation of features sets extracted through individual models. The selection of the method is motivated by its proven performance in literature [25]. DCA based fusion mechanism aims to mitigate the drawbacks of feature-level fusion using Canonical Correlation Analysis (CCA) [26]. The basic insight of the DCA based fusion is to maximize the pair-wise correlations among the feature sets, eliminate the between-class correlation as well as restrict the correlations to be within classes [25].

After transformation of the feature sets with DCA, the transformed features sets are combined in both ways i.e., concatenation and summation operations as recommended in [25], [26]. It is important to mention that DCA based transformation could be applied to only two sets of features at a time. To adopt the method to our four models,
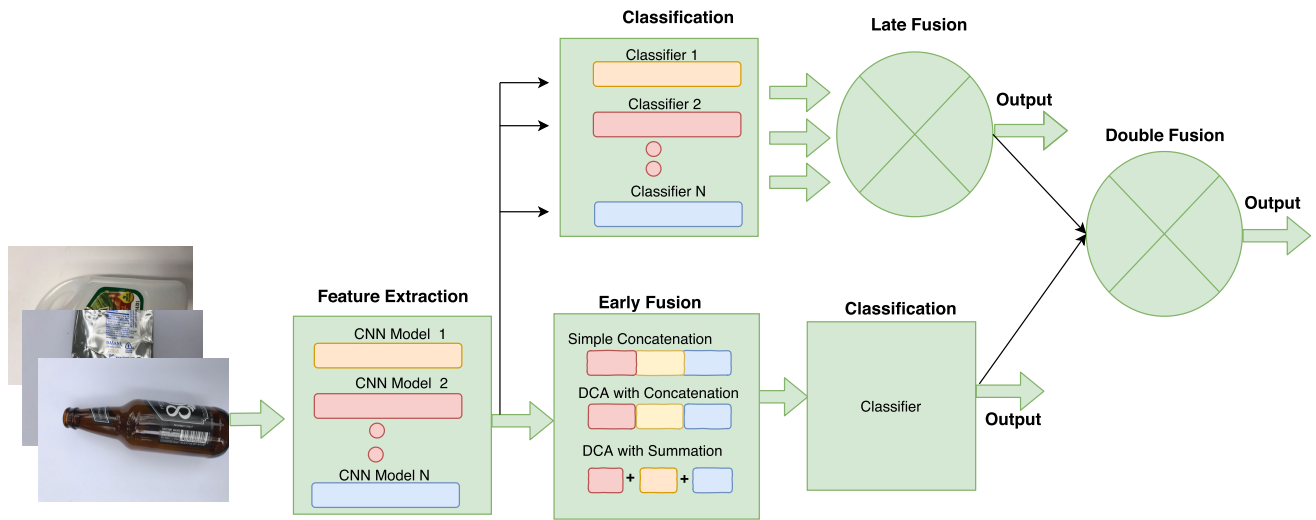
---

[3] https://caffe.berkeleyvision.org/installation.html
[4] https://www.mathworks.com/help/stats/fitcecoc.html

FIGURE 1. Block diagram of the proposed methodology.

TABLE 1. Summary of the deep architectures used for features extraction in this work in terms of total number of layers, parameters, feature extraction layer and the length of the feature vectors obtained with them.

Model	Total Layers	Total Parameters	Features Extraction layer	Feature Vector Length
AlexNet [21]	8	60	FC-7	4096
GoogleNet [9]	22	102	Cls-FC-2	1000
VGGNet [22]	19	144	FC-7	4096
ResNet-101 [23]	101	44.5	FC-1000	1000

we divided our features pool into two groups; one containing AlexNet and VggNet and the other covering GoogleNet and ResNet based on the length of their feature vectors. However, the feature sets don't need to be of the same dimension. DCA based transformation could also be applied to feature sets with different dimensions.

2) LATE FUSION SCHEMES
Similar to early fusion, several techniques have been employed for the late fusion of the scores obtained with classifiers trained on features extracted with the individual models. For late fusion, we mainly explored and used four methods, namely (i) simple averaging as a baseline method, (ii) Particle Swarm Optimization (PSO) (iii) Genetic Algorithms (GA) and (iv) Induced Ordered Weighted Averaging (IOWA) based fusion. In the first method, classification scores are simply averaged by treating them equally, and a class with the highest average score has been selected as the final label of the image. On the other hand, PSO, GA, and IOWA based methods assign merit-based weights to the models in the fusion process. It is important to mention that PSO and GA based methods need to learn the weights while IOWA based fusion does not require a validation set to learn weights to be assigned to the models. A detailed description of all of the methods has been provided in the next subsection. After weights optimization with PSO and GA algorithms, the deep models are combined using Equ. 1. Here, $S_{combined}$ represent the final score, $w(n)$ is the weight assigned to n^{th}

model and p_n represents the probability vectors obtained with the classifier trained on the features extracted with it. Subsequently, an image is labeled with the waste category with the highest combined score.

$$S_{combined} = w(1) * p_1 + w(2) * p_2 + \ldots + w(n) * p_n \quad (1)$$

a: PSO BASED WEIGHTS OPTIMIZATION
The Particle Swarm based optimization of the weights assigned to the classifiers trained on features extracted with the deep models is motivated by the proven performance of the method in literature [12], [27]. PSO, which is inspired by the social behavior of birds flocking, starts with a randomly selected population of potential solutions and tries to optimize/improve the candidate solutions concerning a given criteria provided in the fitness function, iteratively. To adopt, the method in our waste classification task, we treat each combination of the weights to be assigned to the models as a candidate solution while our objective/fitness function is based on the accumulative error ($error_{acc}$) computed by Equ. 2. Here A_{acc} represents the accumulative accuracy computed on the validation set using Equ. 3 where p_n represents the probabilities obtained through n^{th} model on the validation set while $x(n)$ represents the value (weight) to be used and optimized for the nth model.

$$error_{acc} = 1 - A_{acc} \quad (2)$$
$$A_{acc} = x(1) * p_1 + x(2) * p_2 + \ldots + x(n) * p_n \quad (3)$$

b: GA BASED MODELING THE DEEP MODELS

Our second merit-based weights optimization/assignment technique is based on GAs [28]. Similar to PSO, in genetic modeling of the deep models, an initial population/generation (initial set of weights) is generated randomly. The method then seeks for optimal solution (set of weights to be assigned to the models) concerning a criteria given in the fitness function, iteratively, over a certain number of generations or until a sufficient level of fitness. In GA based optimization, fitness function and chromosomes are the key components while crossover and mutation are the two basic operators controlling the optimization process. The Crossover tries to pull the population towards local minimum/maximum and the later tries to discover better minimum/maximum space among the local minimum/maximum spaces. In our case, the fitness function is based on accumulative error computed on the validation set using Equ. 1 and Equ. 2. On the other hand, the deep models represent the chromosomes, which is a set of parameters defining the underlying solution, the genetic algorithm trying to solve/optimize.

c: INDUCED ORDERED WEIGHTED AVERAGING OPERATORS BASED LATE FUSION

Our third merit-based fusion method is inspired by non-linear operators, namely IOWA [29], where the weights to be assigned to the classifiers trained on the individual deep models are directly inferred from the posterior probabilities produced by the classifiers. For instance, to combine classification scores obtained with N deep models on the dataset with M classes ($M = 6$ in our case), the method forms a NxM matrix with each entry representing posterior probabilities of a test sample concerning all classes in the dataset. Next, suppose p_i represents the score array obtained by the ith classifier (i.e., a row in the matrix). The fusion strategy aims at gathering an ensemble of pairs $[p_i, o_i]$, where p_i and o_i represent the argument and the corresponding order-inducing value, respectively. The order inducing value associated with the array p_i, containing the posterior probabilities obtained with the ith classifier, quantifies the confidence level of the classifier, which is, in this work, measured by computing the standard deviation of the highest values in the probabilities array (i.e., p_i). Finally, the final decision is made from a weighted sum of the reordered posterior probability vectors, is given by Equ. 4:

$$F(p_i, o_i) = \frac{1}{N} \sum_{i=1}^{N} w_i s_i \qquad (4)$$

where $W = [w_i, w_2, \ldots, w_N]$ and $S = [s_1, s_2, \ldots, s_N]$ represent the associated weights vector and the reordered probabilities vector on the basis of the inducing values, respectively. In literature, different methods have been proposed to obtain the weight vector, such as argument-based, prioritization aggregation, optimization and learning from data [30], [31]. In this work, we used the prioritization

TABLE 2. Class-wise distribution of the images in the dataset.

Category	No. of Images	Category	No. of Images
Cardboard	403	Glass	501
Metal	410	Paper	594
Plastic	482	Trash	137

aggregation method [30] based on its proven performances in [32], [33].

3) DOUBLE FUSION

In the double fusion method, we combine the classification scores obtained as a result of best early and late fusion methods, further, in a late fusion method where PSO and GA based optimization methods are used to measure the contributions of both fusion schemes. It is important to mention that in the double fusion method our fitness criteria are based on the accumulative error computed on the validation set using Equ. 2 where accumulative accuracy (A_{acc}) is computed using Equ. 5. Here $p_{earlyfusion}$ and $p_{latefusion}$ represent the posterior probabilities obtained with early and late fusion on the validation set, respectively.

$$A_{acc} = x(1) * p_{earlyfusion} + x(2) * p_{latefusion} \qquad (5)$$

IV. EXPERIMENTS AND RESULTS

A. DATASET

The evaluation of the proposed methods has been carried on a benchmark large-scale dataset [15]. The dataset we used for the evaluations of the proposed methods, which is the most commonly used dataset in literature, is based on the assumption that the waste materials are already sorted. The dataset contains a total of 2,527 images from six different categories of waste, namely cardboard, glass, metal, paper, plastic, and trash. Table 2 provides class-wise distribution of the images in the dataset. Moreover, the dataset has been provided in separate training and test sets. The images are captured with three different devices, namely Apple iPhone 7 Plus, Apple iPhone 5S, and Apple iPhone SE. Each image contains a single object. The dataset aims at the classification of the waste material instead of the detection of the waste items. During data acquisition, the object of interest is kept on a white post-board. Figure 2 provides some sample images from the dataset.

B. EXPERIMENTAL SETUP

The objective of our analysis is manifold. On one hand, we want to assess the performance of the individual deep models in waste classification tasks. On the other hand, we want to analyze how the performance is affected when multiple deep models are used jointly. We also want to evaluate different fusion schemes from both early and late fusion families. To attain these goals, we performed different experiments. We started with the analysis of the individual deep models. Then, we investigate the performances of the deep models jointly in two early and three different late fusion techniques.
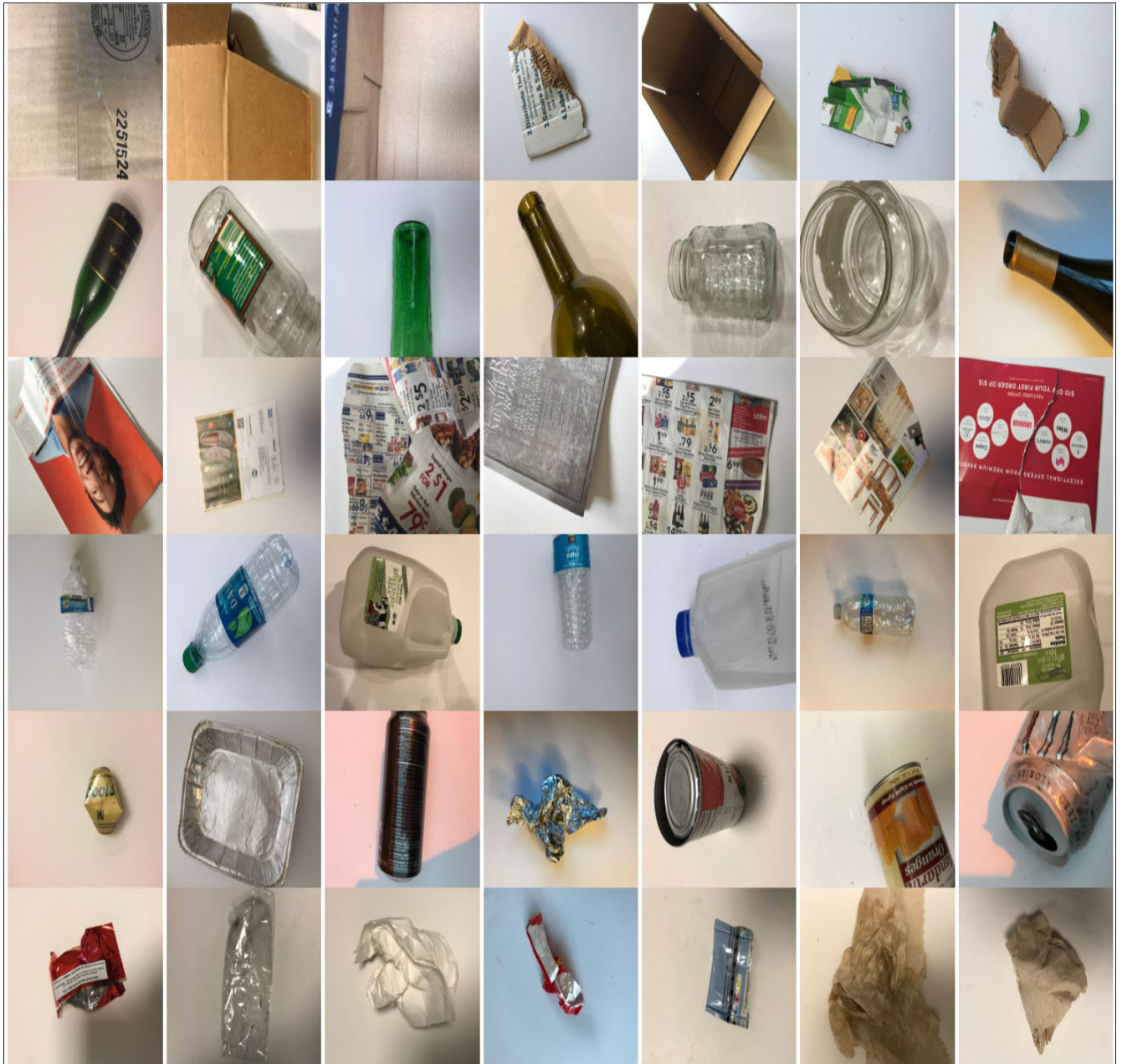
FIGURE 2. Sample images from the dataset.

Finally, we combined the results of early and late fusion experiments in our final experiment termed as double fusion.

During the experiments, the same experimental setup has been used throughout the experimentation process. For instance, the same validation set has been used for two of our late fusion methods, namely PSO and GA, that require to learn/optimize the weights. IOWA is learning free method and does not require the validation set. Similarly, the same parameters have been used for all the algorithms. For instance, default parameters have been used for SVM (Fit Multiclass model from Mathworks) during all experiments. Similarly, PSO and GA are evaluated under same conditions (i.e., max number of iteration is 200*number of variables (in our case number of variables is 4), Max Stall Iterations = 20, Min Neighbors Fraction = 0.25 and Swarm Size = min(100,10*number of variables)) in all experiments.

C. EXPERIMENTAL RESULTS

Table 4 provides the experimental results of our first experiment where we evaluate the individual deep models in terms of accuracy, average precision, recall, and F1-score. Overall, better results have been obtained with ResNet-101. Significant variation can be observed in the performances of the models, generally, better results are obtained with deeper
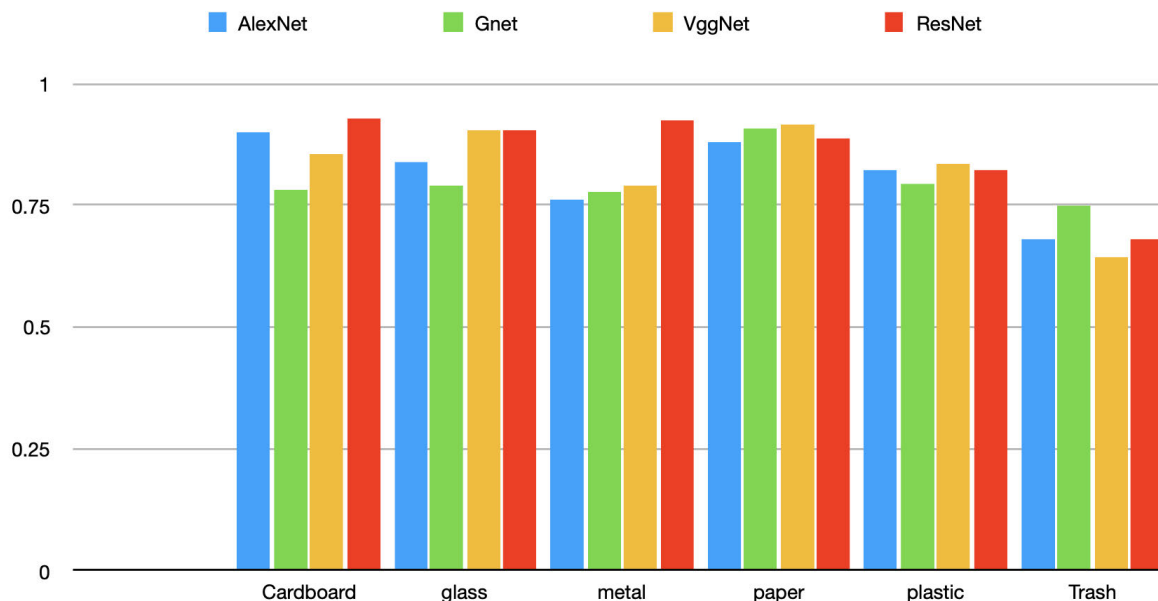
FIGURE 3. Evaluations of the deep models in terms of accuracy per class.

TABLE 3. Confusion matrix of the best individual deep model (ResNet).

Cardboard	Cardboard	Glass	Metal	Paper	Plastic	Trash
	0.92	0	0	0.07	0	0
Glass	0	**0.90**	0.03	0	0.06	0
Metal	0	0.02	**0.92**	0.01	0.02	0
Paper	0.04	0	0.03	**0.88**	0.009	0.01
Plastic	0	0.10	0.05	0.01	**0.82**	0
Trash	0	0	0.07	0.10	0.14	**0.67**

TABLE 4. Evaluation of the individual deep models employed as feature descriptors in waste classification.

Model	Accuracy	Avg Precision	Avg Recall	F1-score
AlexNet	83.29	81.30	82.91	82.10
GNet	81.41	79.99	82.27	81.12
VggNet-19	85.17	82.36	83.99	83.17
ResNet-101	87.76	85.70	87.94	86.81

models except for GNet, which surprisingly achieved the least accuracy, precision, and recall. To analyze the performances of the models on the individual categories, in Figure 3 we provide the evaluations of the models in terms of accuracy per class. Generally higher accuracy has been observed in most of the classes except Trash where the scores for all the models are significantly lower compared to the rest of the classes. One of the possible reasons is the lower number of samples in the class. Another reason could be the high intra- and low inter-class variations concerning other classes, which has also been confirmed by manual inspection of the trash related images. The confusion matrix of the best individual deep model (ResNet) provided in Table 3 also depicts the confusion among certain classes. As can be seen, Cardboard is mostly confused with Paper, Glass images are mostly mis-classified as Metal and Plastic. Similarly, metal is sometimes confused with Plastic, glass, and paper while trash items are usually is-leaded as Metal and Plastic.

Another interesting observation is the comparisons of the models in terms of accuracy per class. As expected, ResNet-101 has outperformed its counterparts in most of the waste classes. However, surprisingly better results have been observed for GNet on trash related images compared to the

other models, which shows that the model is more biased towards the class.

To better analyze the variation in the performances of the models, in Figure 4, we provide a standard deviation of the accuracy obtained with models per class. Though, overall higher variations have been reported on the Metal class where the main contributor in the standard deviation is ResNet-101 model while the rest of the models have similar results. On the other hand, in the case of Cardboard and Glass classes, significant variations can be observed among the performances of the model. As evident from the figure, overall, the models employed in this work responded differently to each class, which motivates our second experiment where we aim to jointly utilize the models using several fusion methods for acquiring a diversified and complementary feature set.

Table 5 provides the evaluations of the fusion methods used in this work. In total seven methods have been evaluated in this work. Several observations have been made during the evaluations. Firstly, a clear advantage can be observed for employing multiple deep models over the individual models. This demonstrates that the different deep models show diverse and complementary image characterization capabilities, and their fusion takes into account such diversity. Secondly, no significant difference has been observed for
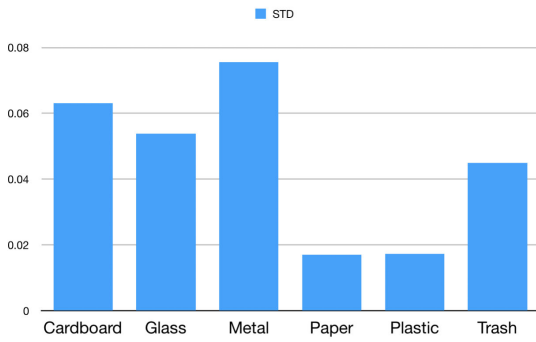
FIGURE 4. Standard deviation of the accuracy obtained with the deep models per class.

TABLE 5. Evaluation of the fusion techniques employed in this work..

Fusion Method	Accuracy	Avg Precision	Avg Recall	F1-score
DCA (concatenation)	88.70	87.10	87.33	87.21
DCA (Summation)	88.47	86.61	87.09	86.85
Simple concatenation	90.82	89.48	90.42	89.94
Late Fusion (Equal Weights)	90.58	88.04	90.50	89.25
Late Fusion (IOWA)	91.05	90.10	89.95	90.02
Late Fusion (PSO)	92.23	90.68	90.87	90.78
Late Fusion (GA)	91.05	88.49	89.90	89.19
Double Fusion (PSO)	94.58	93.41	94.27	93.84
Double Fusion (GA)	94.11	92.25	93.76	93.00

the simple late fusion (equal weights) and early fusion (simple concatenation) methods. However, significant improvements have been noticed for merit-based late fusion methods (PSO, IOWA, and GA) over simply averaging the classification scores with equal weights for all models. As shown in Figure 4, the performances of the models vary, thus treating them equally in fusion is not an optimal way. The merit-based methods (PSO, IOWA, and GA) ensure to maximize the contribution of the better performing models in the final decision by assigning them higher weights. On the other hand, very surprising results have been observed in the case of early fusion where slightly least accuracy has been obtained with DCA based fusion over the simple concatenation.

As can be seen in Table 5, early and late fusion methods respond differently with a reasonable variation in the performances. This variation in the performance of the fusion methods provides the basis for our double fusion method where we combine the capabilities of early and late fusion schemes by averaging their scores using the better performing late fusion methods (PSO and GA), which outperformed all the methods proposed in the work. A significant improvement of 2.35% has been observed for double fusion with PSO over the best single fusion method.

We also provide the confusion matrix of the best performing method (double fusion with PSO) in Table 6 to better highlight the inter-class variability and the confusion among the classes. As can be seen in the table, several classes have been confused with each other by the classifier. For instance, cardboard has been mostly confused with paper while glass-related images are sometimes mis-classified as metal and Plastic. Similarly Metal related images have been sometimes

TABLE 6. Confusion matrix of the double fusion method with PSO.

Cardboard	Cardboard	Glass	Metal	Paper	Plastic	Trash
	0.927	0	0	0.072	0	0
Glass	0	**0.962**	0.012	0	0.024	0
Metal	0.014	0.029	**0.940**	0	0.014	0
Paper	0.018	0	0	**0.971**	0	0.009
Plastic	0	0.027	0.013	0	**0.945**	0.013
Trash	0	0	0.071	0	0.071	**0.857**

TABLE 7. Comparisons against state-of-the-art methods.

Fusion Method	Accuracy
Baseline [15]	87.00
Ruiz et al. [34]	88.60
RecycleNet [16]	81.00
Aral et al [35]	91.00
PSO based Late Fusion (Ours)	92.23
GA based Late Fusion (Ours)	91.05
Early Fusion (simple concatenation-Ours)	90.58
Double Fusion (GA)	**94.11**
Double Fusion (PSO)	**94.58**

confused with cardboard, glass, and plastic. Moreover, paper and trash are sometimes confused with each other.

We also provide a comparison of our double fusion methods against state-of-the-art methods for waste classification in Table 7. It is important to mention that for comparison purposes, we just took our better-performing methods only. As can be seen, our double fusion methods with PSO and GA obtained around 3.5% and 3.1% improvements over the state-of-the-art method, respectively. The significant improvement in the performance shows the effectiveness of the proposed double fusion method.

V. CONCLUSION AND FUTURE WORK

The paper addressed the problem of waste classification from two complementary perspectives. On one side, it is demonstrated that the fusion of multiple deep models outperform the individual models by jointly exploiting the learning capabilities of individual deep models. On the other side, we conduct a comprehensive analysis of the state-of-the-art deep models. We then demonstrate that better waste classification performances can be achieved by optimally combining the deep models through appropriate fusion methods. All the fusion methods employed in this work are based on the assumption that the state-of-the-art models respond differently by extracting diverse and complementary image features/characteristics, and the fusion of these diverse and complementary image features could result in an improvement of the waste classification frameworks. Based on our experiments, we conclude that combining multiple deep models for waste classification is a more reliable method. Moreover, the fusion of early and late fusion methods outperforms the best individual fusion methods.

In future a new dataset covering a higher number of waste categories images with different levels of complexity will be collected.

ACKNOWLEDGMENT

REFERENCES
[1] J. P. Parfitt, A. A. Lovett, and G. Sünnenberg, "A classification of local authority waste collection and recycling strategies in england and wales," *Resour., Conservation Recycling*, vol. 32, nos. 3–4, pp. 239–257, Jul. 2001.

[2] G. Davidson, "Waste management practices: Literature review," in *Dalhousie University, Office of Sustainability*, 2011, pp. 1–59.

[3] K. Sander, S. Schilling, H. Lskow, J. Gonser, A. Schwedtje, and V. Kchen, "Review of the European list of waste," Ökopol GmbH Cooperation ARGUS GmbH, Hamburg, Germany, Tech. Rep., 2008.

[4] Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong, "Multilayer hybrid deep-learning method for waste classification and recycling," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–9, Nov. 2018.

[5] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A novel framework for trash classification using deep transfer learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019.

[6] O. Awe, R. Mengistu, and V. Sreedhar, "Final report: Smart trash net: Waste localization and classification," Tech. Rep., 2017.

[7] U. Ozkaya and L. Seyfi, "Fine-tuning models comparisons on garbage classification for recyclability," 2019, *arXiv:1908.04393*. [Online]. Available: http://arxiv.org/abs/1908.04393

[8] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[10] K. Pogorelov, O. Ostroukhova, M. Jeppsson, H. Espeland, C. Griwodz, T. de Lange, D. Johansen, M. Riegler, and P. Halvorsen, "Deep learning and hand-crafted feature based approaches for polyp detection in medical videos," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2018, pp. 381–386.

[11] K. Ahmad and N. Conci, "How deep features have improved event recognition in multimedia: A survey," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 2, p. 39, 2019.

[12] K. Ahmad, M. L. Mekhalfi, N. Conci, F. Melgani, and F. De Natale, "Ensemble of deep models for event recognition," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 2, p. 51, 2018.

[13] J. Donovan, "Auto-trash sorts garbage automatically at the techcrunch disrupt hackathon," Techcrunch Disrupt Hackaton, San Francisco, CA, USA, Tech. Rep. Disrupt SF 2016, 2016.

[14] G. Mittal, K. B. Yagnik, M. Garg, and N. C. Krishnan, "Spot-Garbage: Smartphone app to detect garbage using deep learning," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2016, pp. 940–945.

[15] *Classification of Trash for Recyclability Status*. Accessed: Jan. 6, 2020. [Online]. Available: http://cs229.stanford.edu/proj2016/poster/ThungYang-ClassificationOfTra%shForRecyclabilityStatus-poster.pdf

[16] C. Bircanoglu, M. Atay, F. Beser, O. Genc, and M. A. Kizrak, "RecycleNet: Intelligent waste sorting using deep neural networks," in *Proc. Innov. Intell. Syst. Appl. (INISTA)*, Jul. 2018, pp. 1–7.

[17] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*. [Online]. Available: http://arxiv.org/abs/1404.1869

[18] J. Sousa, A. Rebelo, and J. S. Cardoso, "Automation of waste sorting with deep learning," in *Proc. 15th Workshop de Visão Computacional (WVC)*, Sep. 2019, pp. 43–48.

[19] Z. Wang, B. Peng, Y. Huang, and G. Sun, "Classification for plastic bottles recycling based on image recognition," *Waste Manage.*, vol. 88, pp. 170–181, Apr. 2019.

[20] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 239–246.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[25] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis for feature level fusion with application to multimodal biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1866–1870.

[26] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng, and D.-S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recognit.*, vol. 38, no. 12, pp. 2437–2448, Dec. 2005.

[27] Y. del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley, "Particle swarm optimization: Basic concepts, variants and applications in power systems," *IEEE Trans. Evol. Comput.*, vol. 12, no. 2, pp. 171–195, Apr. 2008.

[28] L. Scrucca, "GA: A package for genetic algorithms in R," *J. Statist. Softw.*, vol. 53, no. 4, pp. 1–37, 2013.

[29] R. R. Yager and D. P. Filev, "Induced ordered weighted averaging operators," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 29, no. 2, pp. 141–150, Apr. 1999.

[30] R. R. Yager, "On prioritized multiple-criteria aggregation," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 5, pp. 1297–1305, Oct. 2012.

[31] X. Liu and S. Yu, "On the stress function-based OWA determination method with optimization criteria," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 1, pp. 246–257, Feb. 2012.

[32] H. AlHichri, Y. Bazi, N. Alajlan, F. Melgani, S. Malek, and R. R. Yager, "A novel fusion approach based on induced ordered weighted averaging operators for chemometric data analysis," *J. Chemometrics*, vol. 27, no. 12, pp. 447–456, Dec. 2013.

[33] K. Ahmad, M. L. Mekhalfi, N. Conci, G. Boato, F. Melgani, and F. G. B. De Natale, "A pool of deep models for event recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2886–2890.

[34] V. Ruiz, A. Sánchez, J. F. Vélez, and B. Raducanu, "Automatic image-based waste classification," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.* Almería, Spain: Springer, 2019, pp. 422–431.

[35] R. A. Aral, S. R. Keskin, M. Kaya, and M. Haciomeroglu, "Classification of TrashNet dataset based on deep learning models," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2058–2062.

KASHIF AHMAD received the bachelor's and master's degrees from the University of Engineering and Technology, Peshawar, Pakistan, in 2010 and 2013, respectively, and the Ph.D. degree from the University of Trento, Italy, in 2017. He worked with the Multimedia lab, DISI, University of Trento. He also worked as a Postdoctoral Researcher with ADAPT, Trinity College, Dublin, Ireland. He is currently working as a Research Fellow with the Division of Computer Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. He has authored and coauthored more than 30 journal and conference publications. His research interests include multimedia analysis, computer vision, machine learning, and signal processing.

KHALIL KHAN received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2007 and 2012, respectively, and the Ph.D. degree from the Signals and Communication laboratory, University of Brescia, Italy, in 2016. He is currently working as an Assistant Professor with the Department of Electrical Engineering, The University of Azad Jammu & Kashmir, Muzafarabad, Pakistan. His research interests include a wide range of topics within digital image processing, machine learning, and computer vision.

ALA AL-FUQAHA (Senior Member, IEEE) received Ph.D. degree in computer engineering and networking from the University of Missouri–Kansas City, Kansas City, MO, USA, in 2004. He is currently a Professor with Hamad Bin Khalifa University (HBKU). His research interests include the use of machine learning in general and deep learning in particular in support of the data-driven and self-driven management of large-scale deployments of the IoT and smart city infrastructure and services, wireless vehicular networks (VANETs), cooperation and spectrum access etiquette in cognitive radio networks, and management and planning of software defined networks (SDN). He is an ABET Program Evaluator (PEV). He has served as the Chair, the Co-Chair, and a technical program committee member for multiple international conferences, including IEEE VTC, IEEE Globecom, IEEE ICC, and IWCMC. He also serves on editorial boards for multiple journals, including the IEEE COMMUNICATIONS LETTER and IEEE NETWORK MAGAZINE.

• • •