Automatic Image-Based Waste Classification

Victoria Ruiz[1], Ángel Sánchez[1], José F. Vélez[1], and Bogdan Raducanu[2]

(1) ETSII - URJC, 28933 Móstoles (Madrid), Spain
{victoria.ruiz.parrado,angel.sanchez,jose.velez}@urjc.es
(2) CVC - UAB, 08193 Cerdanyola (Barcelona), Spain
{bogdan@cvc.uab.es}

Abstract. The management of solid waste in large urban environments has become a complex problem due to increasing amount of waste generated every day by citizens and companies. Current Computer Vision and Deep Learning techniques can help in the automatic detection and classification of waste types for further recycling tasks. In this work, we use the TrashNet dataset to train and compare different deep learning architectures for automatic classification of garbage types. In particular, several Convolutional Neural Networks (CNN) architectures were compared: VGG, Inception and ResNet. The best classification results were obtained using a combined Inception-ResNet model that achieved 88.6% of accuracy. These are the best results obtained with the considered dataset.

Keywords: Computer Vision, Deep learning, Convolutional Neural Networks, Waste Classification

1 Introduction

Waste collection and recycling are essential services for modern cities, specially for the big ones. Due to a decrease of available natural resources and to environmental problems produced by the increasing amount of generated garbage, there is a need for recycling to reduce pollution and health problems for citizens. The average European generates 517 kilos of garbage per year, of which just a small percentage is recycled [1]. According to Environmental Protection Agency, 75% of waste produced by American people is recyclable, but actually only 30% is recycled. Currently, most of the garbage segregation process is done manually which creates many health problems for the workers, is time-consuming and also requires financial taxes from citizens [2]. Moreover, this waste separation must be done as soon as possible in order to reduce the contamination of waste by other materials [3].

Waste separation and recycling is necessary for a sustainable society. Currently, the application of ICT (e.g. using technologies and devices such as smart sensors, cloud platforms or Internet of Things) to smart cities in automatic garbage classification tasks can significantly improve the efficiency of these processes [1]. This classification can be made by the type of garbage [4], the biodegradable nature of the waste [2], or other aspects [5]. On the other hand, anti-littering organizations and cities governments worldwide are assessing urban cleanliness by means of human audits [6]. Waste locating and quantification is an important step for improving cleanliness of cities, which could become health problem in overpopulated countries such as India [5].

These automatic garbage recycling systems can also involve Computer Vision to analyze the images or videos captured in recycling plants to determine which kind of objects are present in mixed waste. Good results in this stage will lead good results in the whole recycling process. Moreover, with recent developments of Machine Learning techniques, specially Deep Learning, very good image-based garbage classification results have been achieved [3].

In this paper, we adopt a supervised approach to effectively classify several types of waste present in images (e.g. glass, paper, cardboard, plastic and so on). For this purpose, we trained and compared several deep classification models to recognize different waste categories present in images of the TrashNet dataset [4]).

The paper is organized as follows. Section 2 reviews image-based systems for waste classification. Section 3 outlines the different deep neural architectures used or supervised classification of waste. Section 4 describes the dataset and waste classification experiments. Finally, Section 5 outlines the conclusions of this study.

2 Previous work

Current Computer Vision systems for waste separation are oriented towards object detection and classification using image analysis techniques. This process could be divided in the following steps:

1. *Segmentation*: It involves separating each type of waste. First, some pre-processings on images are required to remove noise (e.g. Gaussian blur), to enhance contrast (e.g. histogram equalization) or to binarize them (e.g. Otsu algorithm). After that, diverse edge detection methods such as Canny or watershed algorithms can be applied to segment the image into homogeneous regions [1].
2. *Feature extraction*: Before the development of Deep Learning techniques, feature extraction methods (i.e. based on shape, texture or color descriptors) were required to extract useful information from segmented regions and built automatic classification models from these features. For example, statistical moments, Fourier-based, Gabor-based descriptors, Histogram Orient Gradients (HOG) are some of the used methods [1], [5]. Additionally, Principal Component Analysis (PCA) was used to reduce the data dimensionality [5] prior to the classification stage.
3. *Learning and Classification*: Once the features are extracted, a classification model is trained to identify the objects in waste. For example, correlation algorithms [7], K-Nearest Neighbors (KNN) [1] or SVM [4] [3]. From the emergence of Deep Learning, diverse types of deep neural architectures as

AlexNet [4], Faster R-CNN [8] or GoogleNet [6] were also applied in the considered problem. Special neural architectures for this application have been recently built, such as GarbNet [5] or OscarNet [9], which are based on pretrained convolutional neural networks architectures such as AlexNet or VGG-19.

One aspect to consider in classification is image resolution. If images are large, a sliding window can be used [6]. Additionally, when the dataset size is small, data augmentation techniques can be applied as in [4]. Most of proposed systems in the bibliography are focused on localization and classification of waste types. Some of these systems have also been implemented as an Android app, as it is the case of SpotGarbage, developed by Mittal et al. [5].

However, a fair comparison of the accuracy among proposed methods is still difficult because many of them use their own datasets. So, each proposed model can be trained using different waste categories. Table 1 compares some of the current image-based deep learning systems for trash classification. As can be appreciated, some good results have been achieved in recent years. Our goal in this work is to evaluate other deep models that improve current state-of-the-art in garbage classification for the TrashNet dataset.

Author (Year)	Dataset (classes)	Methodology	Accuracy
Briñez et al. (2015) [7]	Own (3)	Correlation algorithm	78.00%
Mittal et al. (2016) [5]	GINI	GarbeNet (based on CNN)	87.69%
Kennedy et al. (2016) [9]	TrashNet (7)	OscarNet (based on VGG-19 pretrained)	88.42%
Sakr et al. (2016) [3]	Own (3)	SVM	94.80%
Sakr et al. (2016) [3]	Own (3)	AlexNet	83.00%
Yang et al. (2016) [4]	TrashNet (5)	SVM with SIFT features	65.00%
Yang et al. (2016) [4]	TrashNet (5)	AlexNet	22.40%
Rad et al. (2017) [6]	Own (25)	Overfeat with GoogleNet	77.35%
Awe et al. (2017) [8]	TrashNet (6)	Augmented data to train R-CNN	68.30%

Table 1. Comparative of recent approaches for garbage classification

3 Deep architectures for supervised waste classification

Many current neural architectures used for supervised classification of images are based on the Convolutional Neural Network (CNN) model. CNN are composed by convolutional layers where neurons are connected through a convolution function instead of a general matrix multiplication so weights are shared rather than being all connected. As a result, spatial patterns which are invariant to translations, rotations, and other transformations, are obtained.

In our experiments, we used several neural architectures, all of them based on convolutional layers. In particular:

1. **VGG**: The VGG architecture was developed for localization and classification tasks on high-resolution images [10]. VGG network is formed by many convolutional layers with increasing depth and with small kernels (i.e. 3×3) in all the convolutional layers. In this work, we have focused in two VGG models:
   (a) <u>VGG-16:</u> In VGG-16 [11], a block of 13 convolution layers and 3 fully-connected layers compose the architecture as follows. One block of two 64-depth convolutional layers with max-pooling, one block of two 128-depth convolutional layers with max-pooling, one block of three 256-depth convolutional layers with max-pooling, two block of three 512-depth convolutional layers with max-pooling, two fully-connected layers with 4096 neurons, one fully-connected layer with as many neurons as classes of the dataset and SoftMax as activation function. Fig. 1(a) shows this architecture.
   (b) <u>VGG-19:</u> VGG 19 [11] is a variation of the previous model. The only difference is that the last three convolutional blocks are formed by 4 convolutional layers instead of 3. Fig. 1(b) shows this architecture.
2. **ResNet**: From Deep Convolutional Networks such as AlexNet or VGG, research has been focused on increasing the depth of the architecture, but the vanishing gradient problem prevented to achieve it. ResNet introduced skip connections to avoid degrading the network performance [12]. As a result, the feature mapping achieved from a convolutional layer is combined with a feature mapping obtained by the previous layer. In our case we have used ResNet-18, which is composed by one block of three 32-depth convolutional layers and four blocks of two convolutional layers with an increasing depths of 64, 128, 256 and 512, respectively. All the convolutional layers have a 3×3 dimensional filters, except for the first two layers which have a 5×5 dimension filters. Finally, on the bottom of the network there are two fully connected or dense layers with 512 and 6 neurons. Fig. 1(c) shows the ResNet-18 architecture.
3. **Inception**: Inception is a deep convolutional neural network which was the state-of-the-art for classification and detection on the Imagenet dataset. Its main contribution is to increase the depth and width of the network while keeping the computational budget constant [13]. The first version of this version is the well known GoogLeNet. In Inception module, the block of convolutional layers are parallel rather than serial as in VGG. This means that, while in the VGG architecture the output of a convolutional layer was the input of the following convolutional layer in a block, in Inception architecture all, or some of the, convolutional layers in a block have the same input and they are concatenated at the end of the block. Fig. 1(d) shows the Inception architecture.
4. **Inception-ResNet** Szegedy et al. [14] combined both Inception and ResNet concepts: residual connections to avoid gradient vanishing and Inception

modules to increase the network by keeping the computational cost. Fig. 1(e) shows the final Inception-ResNet architecture.

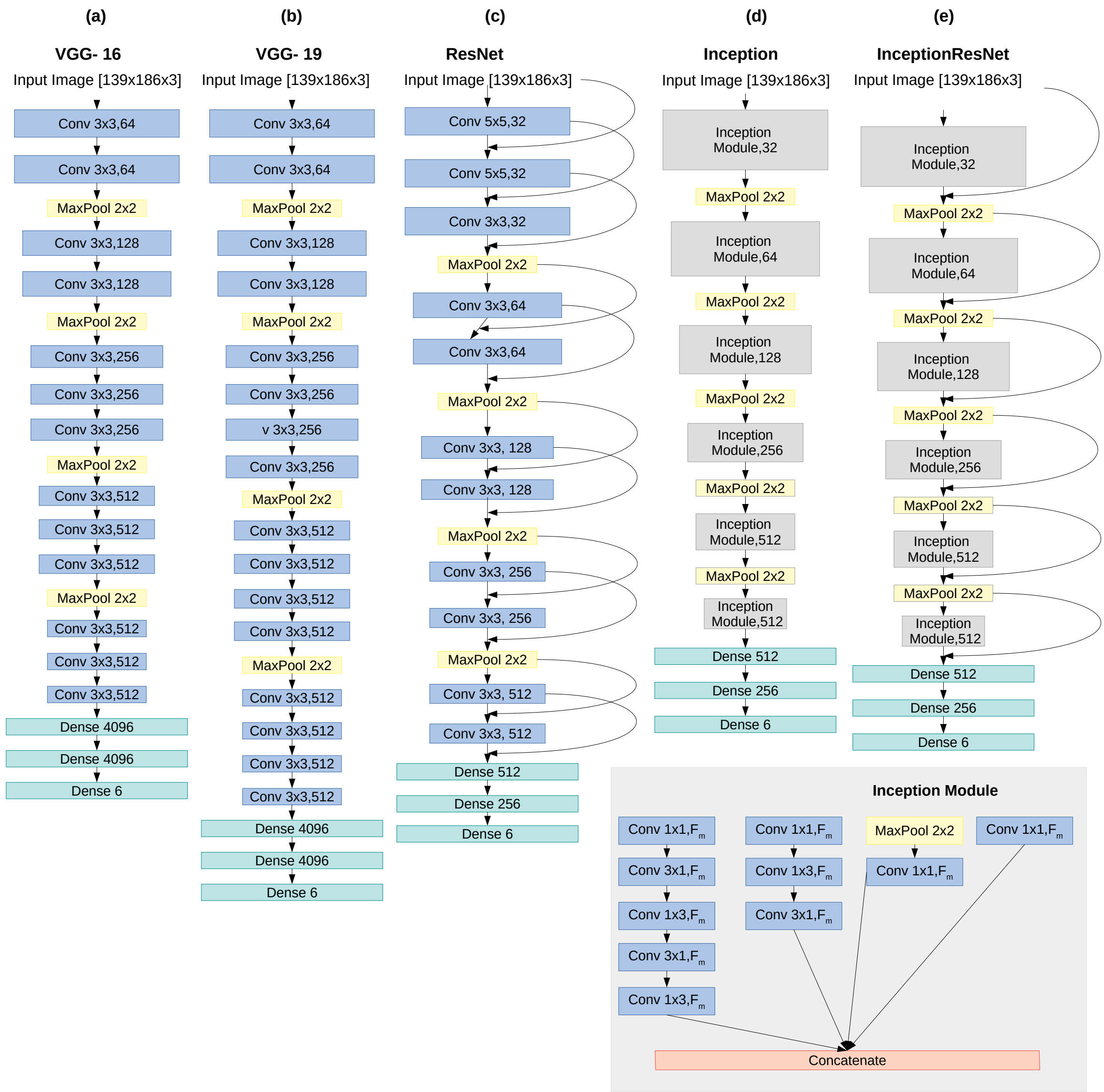

Fig. 1. Additional tested deep models: (a) VGG-16; (b) VGG-19 architectures (c) ResNet; (d) Inception; and (e) Inception-ResNet architectures

4 Classification experiments using TrashNet

In this section, we summarize the dataset used, the pre-processing performed on images and the experiments carried out.

4.1 The TrashNet dataset

The TrashNet dataset [4] was created by Mindy Yang and Gary Thung at Standford University. This dataset contains RGB images of six classes of waste, where in each image only appears one type of garbage. In particular: glass, paper, cardboard, plastic, metal, and general trash, respectively. Currently, the dataset consists of 2,527 images with the following distribution of images per class: 501 of glass, 594 of paper, 403 of cardboard, 482 of plastic, 410 of metal and 137 of general trash, respectively. The images were captured by placing the object on a white posterboard and using sunlight and/or room lighting. All the pictures have been resized down to a spatial resolution of 512×384. Fig. 2 illustrates the six classes present in TrashNet dataset.

As deep neural networks require larger datasets, a common practice is to augment the original collection of original images by applying a set of transformations on each of them (i.e rotations, scalings or brightness corrections, among others).

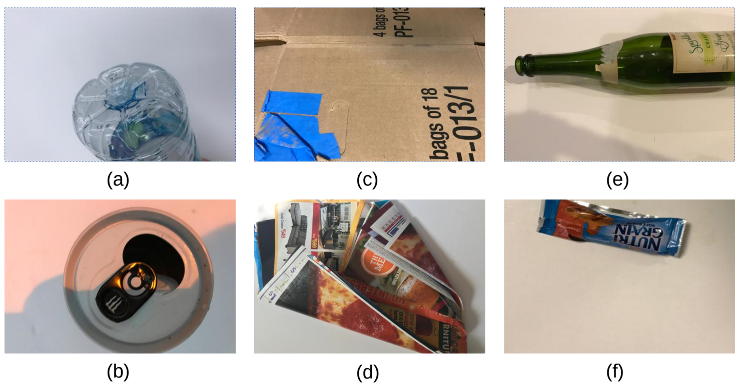

Fig. 2. Sample images of waste classes in TrashNet dataset: (a) plastic; (b) metal; (c) cardboard; (d) paper; (e) glass; and (f) general trash.

4.2 Data pre-processing

Our first goal is to develop a deep learning model which classifies isolated garbage elements using the TrashNet dataset. For this purpose, we tried all the deep convolutional architectures explained previously. However, for all the models we needed to resize the images due to computational problems and to normalize their brightness values between 0 and 1.

Moreover, since we have a small number of images to train our models, data augmentation was used to generate a pseudo-infinite on-demand number of training samples. New images were generated at the same time the model was training, by applying combination of transformations on the original data. Transformations were chosen randomly between: rotations between 0 and 40 degrees, width changes between 0% and 20%, height changes between 0% and 20%, shear between 0% and 20%, zoom between 0% and 20%, and horizontal flips.

4.3 Classification experiments and results

We first randomly partitioned the original collection of images into three subsets: training, validation and test, respectively. All the subsets have the same rate of classes. As the number of images is small, we decided to use 80% of them for training, 10% for validation and the remaining 10% for tests. To achieve more robust results, we adopted a 5-fold cross-validation strategy, by creating randomly 5 training/validation/test sets. Moreover, as it was explained before, the training sample is increased through data augmentation technique. The results given in this subsection correspond to the average of the 5 runs of the test datasets.

The second stage was to configure the parameters of each network. The networks' weights were in all cases initialized randomly. For all of the networks we considered in our experiments, we used a batch size of 16 samples, a Stochastic Gradient Descent (SGD) as optimization algorithm and a learning rate of 0.0002. An early stopping strategy was adopted during training. We kept the model with less validation loss and stopped the training if this result did not improve in 25 epochs time. Moreover, batch normalization layers were introduced at the end of each block of convolutional layers in all the models. The images were resized to 197×283 pixels to train the model.

Table 2 presents a comparative study in terms of mean and standard deviation accuracy results achieved using the five considered deep networks tested. Also, a comparative study of the epochs needed to train the models is shown on this Table. On one hand, best results are achieved by the ResNet model with a 88.66% of accuracy. Moreover, ResNet model is the most stable one since the standard deviation is the smallest. However, the Inception-ResNet model produced similar results. On the other hand, ResNet model is the one which needs less epochs to be trained. We can conclude that the ResNet model is the best by accuracy and speed.

Table 3 compares our best results, achieved by the ResNet model, with other deep learning models applied on waste classification. It is shown that our model wins all the other models, although is quite close to the Kennedy et al. [9] model. However, Kennedy mixed TrashNet and PASCAL data sets, with class 7 (non waste) being the second data set. On the other hand, it is shown in his results that they overfitted the model, achieving good results in the non-waste class (PASCAL data set) but low results for the TrashNet dataset.

model	mean accuracy (in %)	std. dev. accuracy	mean no. epochs	std. dev. no. epochs
VGG-16	76.94	5.75	74.6	37.06
VGG-19	79.32	4.66	76.8	27.35
Inception	87.71	3.36	43.8	16.13
ResNet	88.66	1.28	45.2	5.93
Inception-Resnet	88.34	1.92	55.2	15.27

Table 2. Accuracy of tested deep neural models

Author (Year)	Dataset (classes)	Methodology	Accuracy
Mittal et al. (2016) [5]	GINI	GarbeNet (based on CNN)	87.69%
Kennedy et al. (2016) [9]	TrashNet (7)	OscarNet (based on VGG-19 pretrained)	88.42%
Sakr et al. (2016) [3]	Own (3)	AlexNet	83.00%
Yang et al. (2016) [4]	TrashNet (6)	AlexNet	22.40%
Rad et al. (2017) [6]	Own (25)	Overfeat with GoogleNet	77.35%
Awe et al. (2017) [8]	TrashNet (6)	Augmented data to train R-CNN	68.30%
Ruiz et al. (2019)	TrashNet	Inception-ResNet Model	88.66%

Table 3. Comparative of deep neural network approaches for garbage classification

Finally, Fig. 3 shows the achieved confusion matrices for each of the different deep architectures tested. As we trained five models for each architecture, we show the model with accuracy closer to the average accuracy. We can not determine the class with more accuracy because it depends on the model.

5 Conclusion

In this paper, we have evaluated the use of several CNN architectures for the automatic classification of waste. In our experiments on the TrashNet dataset, best classification results were achieved using a ResNet architecture with 88.66% of average accuracy. Furthermore, we have achieved the best results on the same dataset compared with existing state-of-the-art. In future work, we would like to research the generation of realistic synthetic images with multiple types of garbage, which will be used to train our models, and afterwards test them with real images that combine several classes of wastes.

Acknowledgements

The authors gratefully acknowledge the financial support of the CYTED Network "Ibero-American Thematic Network on ICT Applications for Smart Cities" (Ref: 518RT0559) and also the Spanish MICINN RTI Project (Ref: RTI2018-098019-B-100).

References

1. I. Rodríguez, J. Pérez, and A. Salmador. Intelligent garbage classifier. *International Journal of Interactive Multimedia and Artificial Intelligence*, 1(1):31–36, 2008.
2. S. Sudha, M. Vidhyalakshmi, and K. Pavithra. An automatic classification method for environment, 2016.
3. G.E. Sakr, M. Mokbel, A. Darwich, M.N. Khneisser, and Ali Hadi. Comparing deep learning and support vector machines for autonomous waste sorting. In *Multidisciplinary Conference on Engineering Technology (IMCET), IEEE International*, pages 207–212. IEEE, 2016.
4. M. Yang and G. Thung. Classification of trash for recyclability status. *CS229 Project Report*, 2016, 2016.
5. G. Mittal, K.B. Yagnik, M. Garg, and N.C. Krishnan. Spotgarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 940–945. ACM, 2016.
6. M.S. Rad, A. von Kaenel, A. Droux, F. Tieche, N. Ouerhani, H.K. Ekenel, and J.-P. Thiran. A computer vision system to localize and classify wastes on the streets. In *International Conference on Computer Vision Systems*, pages 195–204. Springer, 2017.
7. J.C. Brinez, A. Rengifo, and M. Escobar. Automatic waste classification using computer vision as an application in colombian high schools. 2015.
8. O. Awe, R. Mengistu, and V. Sreedhar. Smart trash net: Waste localization and classification. 2017.
9. T. Kennedy. Oscarnet: Using transfer learning to classify disposable waste.
10. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
11. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.
12. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
13. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
14. C. Szegedy, S. Ioffe, V. Vanhoucke, and A.A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

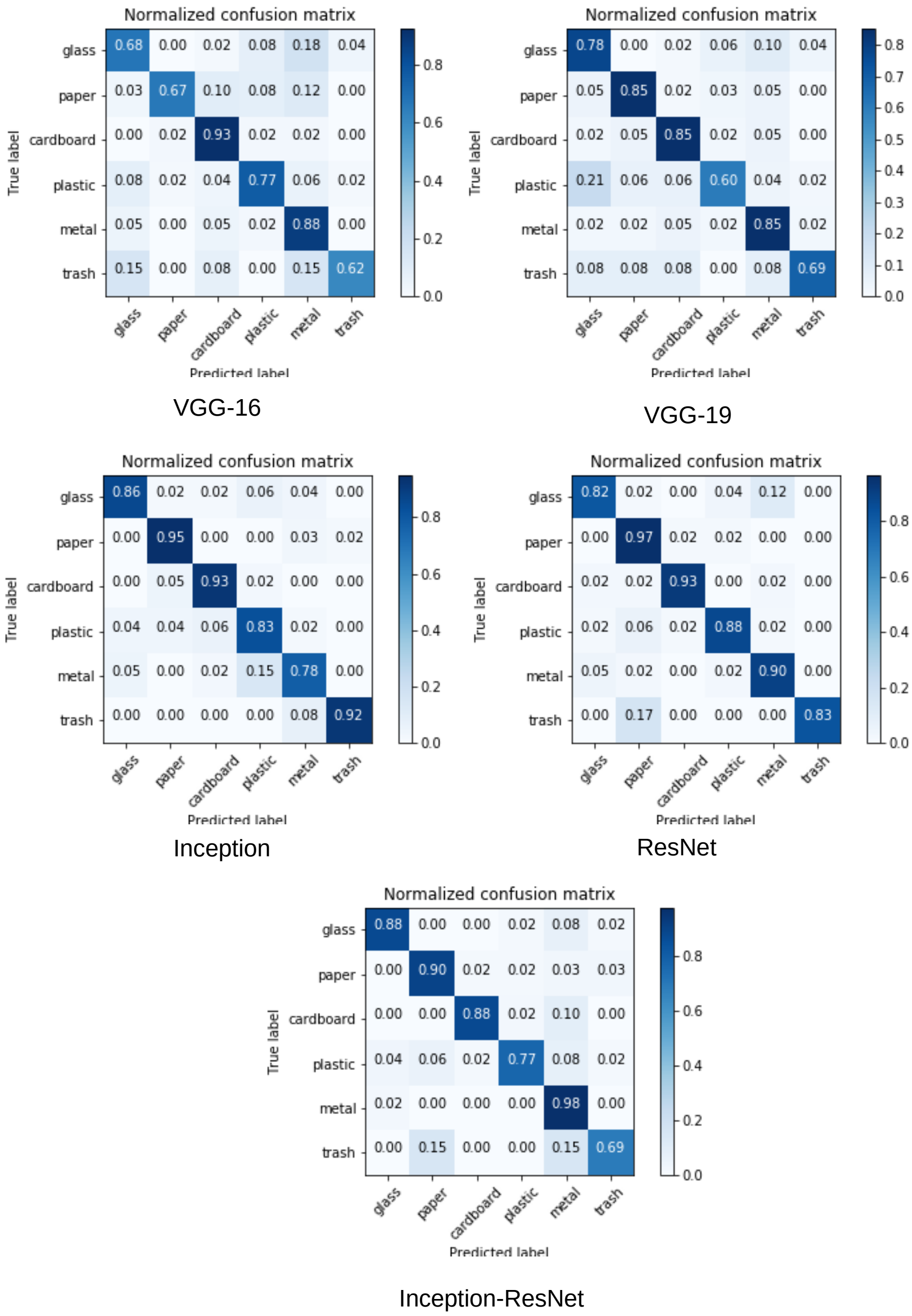

Fig. 3. Comparative of Confusion Matrix