2020 International Russian Automation Conference (RusAutoCon)

Machine Learning Methods for Solving Scrap Metal Classification Task

Nikolai V. Smirnov
Petrozavodsk State University
Petrozavodsk, Russia
nvsmirnov87@gmail.com

Egor I. Rybin
Petrozavodsk State University
Petrozavodsk, Russia
rybin@cs.petrsu.ru

***Abstract*—This paper deals with the task of scrap metal images classification. The authors proposed the method for automation of the cropping scrap metal images from convex quadrangle process and applied this method on railway carriages photographs. The brief description of convolutional neural networks (CNN) and machine learning methods used during the research is given in the paper. The paper presents the results of using various CNN and machine learning methods in the task of classifying images of scrap metal. The algorithm of improving image classification results is proposed. The results of the calculation showed high classification accuracy and allowed to choose the best classifier.**

Keywords—machine learning, convolutional neural network, scrap metal, image classification

I. Introduction

Application of machine learning, deep learning and computer vision methods provide solving a wide range of problems [1–7]. The authors in [8] used multispectral images as input images and proposed using deep learning and machine learning to automate the scrap metal classification process.

One might be mistaken during image recognition process. An error in scrap metal type classifying can cause significant financial losses of a metallurgical plant. Automation of the process using methods of computer vision and machine learning could help the plant to avoid financial losses. A plant employee can apply these methods using photographs or a video stream. The task of this paper is to develop automated solution for classification of scrap metal.

The classification of scrap metal images is a quite complicated task due to the outward similarity of different scrap metal types. Furthermore, the researchers often could have some difficulties such as insufficient number of images in the training set, photographs made from different angles relative to the axis of the railway carriage, low quality of photographs: poor lighting, glare of the sunlight in the photograph.

The paper concentrates on image classification problem and its aim is to predict type of scrap metal based on images of a scrap metal only. Solving this problem requires an additional work to be done to obtain scrap metal images from original photographs. Manual way of cropping the images takes a lot of time and requires human interaction. It necessitates the automation of cropping images in railway carriage photographs. It should be noted that specialized methods for object detection exist, such as, for instance, YOLO [9] or R-CNN [10], whereas, approach proposed in this paper is used for solving wider range of tasks.

II. Used Methods

There is a plenty of different approaches to solving classification tasks. One of the most recent and promising approach is using of the deep convolutional neural networks. Researchers also use other machine learning methods for solving classification tasks.

A. *Convolutional Neural Networks*

Convolutional neural networks used in this paper performed one of the best results in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [11]. ILSVRC is a software contest in which programs compete each other in task of correct and accurate classification and detection of objects and scenes. ILSVRC is a sub-project of ImageNet. ImageNet is a project that collects, annotates and maintains large database of images for using it in computer vision research application. ILSVRC uses 1000 categories while ImageNet itself contain more than 20000 different categories.

The authors used convolutional neural networks with different structure InceptionV3 [12], ResNet50 [13], Xception [14].

1) *InceptionV3*

InceptionV3 is CNN developed by Google. It uses special Inception modules. The Inception module is a small neural network by itself, that consist of multiple convolutional layers with different kernel sizes.

Trained on ILSVRC this neural network was able to achieve accuracy of 0.779 in Top-1 and 0.937 in Top-5.

2) *ResNet50*

ResNet50 is CNN that uses residual layers. It simplifies training process and allows creation of deeper neural networks in comparison to the previous ones. ResNet50 was developed by Microsoft.

978-1-7281-6130-3/20/$31.00 ©2020 IEEE

Trained on ILSVRC this neural networks have accuracy of 0.749 in Top-1 and 0.921 In Top-5.

3) Xception

Xception continues improving of Inception modules by trying to add residual layers and combine work that was done in researches of InceptionV3 and ResNet neural networks.

Trained on ILSVRC Xception achieved accuracy of 0.790 in Top-1 and 0.945 in Top-5, which is better than InceptionV3 or ResNet50 results.

B. Other machine learning methods

In addition to deep learning methods other machine learning methods was used such as: Logistic Regression [*15*], Random Forest [*16*], Gradient Boosting [*17*] and Voting. The methods mentioned above are able to perform better results on some datasets, nevertheless, they are generally not as accurate as CNN, described in previous part. The main advantages of using these methods are the time saving during the training and lesser task computational complexity.

1) Logistic Regression

Logistic Regression is a linear classifier.

2) Random Forest

Random Forest is an ensemble of decision trees trained on different sub-samples from original dataset. Each tree is determined to learn some random subset of features of image.

3) Gradient Boosting

Gradient boosting is one of the boosting algorithms. Idea of boosting is to combine multiple weak learners into one strong learner.

4) Voting

Voting is a simple combination of previous three methods. Each of the methods classifies the image. Voting method choose class that was predicted by most of classifiers.

C. Local Binary Patterns

Convolutional neural networks operate with an image as it is: taking matrix of values representing pixels of the original image to the input.

Other machine learning methods require an additional work to be done to convert images into the feature vectors. Local Binary Patterns method [18] (LBP) was used to create such feature vector. LBP is a mere and effective tool in computer vision, which together with a computational simplicity makes it a commonly used approach for feature extraction.

III. Dataset

A. Original dataset

The dataset is set of photographs of railway carriages filled with different kind of scrap metal (Fig. 1). The authors used only those parts of railway carriage photographs that contain scrap metal, the rest part of photographs was cut out.

Fig. 1. Railway carriages.

B. Preparing dataset

Applying neural networks for image classification task requires certain size of images in dataset. The authors chose 224 by 224 pixels by 3 color channels as an actual image size. Such sub-images were cropped out from original photographs of railway carriages. Obviously, there are many ways of creating sub-images from original, therefore, the following restrictions were created:

- Each of sub-image must be square.
- Given task is to classify scrap metal, thus, each sub-image must be taken from inside of a railway carriage.
- Either size of sub-images must be specified or they must take as much space as possible within railway carriage borders. Sub-images must be scaled to the required size after they are cut out.
- None of Sub-images overlap should exceed 25 percent between each other, preferably, they should not overlap each other at all.
- The sub-images perimeter must be equal.

The authors assume that railway carriage borders are already set either manually or by some other algorithm. Therefore, mathematically this task can be described as: there exist some convex quadrangle (railway carriage) and it is needed to inscribe arbitrary number of squares inside (sub-images). For this purpose Z3 (Satisfiability modulo theories solver) [19] was used. Z3 is not only a theories solver, but it also has abilities for optimization. A larger sub-image will contain more information therefore we maximized the size of sub-images when cropping them from photographs. Nevertheless, Z3 can’t fully optimize the nonlinear equation, and therefore in this paper Manhattan distance is used to calculate the distance between points and the perimeter of the square is maximized instead of the area maximizing. Result of Z3 applying is coordinates of vertices of inscribed squares that was used to crop corresponded parts of the original photograph.

The developed method has wider applying. The method allows finding required size squares in any convex quadrilateral. There is an example of the method applying for finding squares in a convex quadrilateral inside of railways carriage with random contents (Fig. 2). The convex quadrilateral is marked with a green line. The found squares are marked with a white line.

Fig. 2. Marked railway carriage.

C. Received datasets

After all preprocessing and cropping images the authors constructed two sub-datasets which consist only from scrap metal images:

- 9 classes (Fig. 3 (a–i)), each one has 120 images.
- 4 classes (Fig. 3 (b, c, e, h)), each one has 200 images.

IV. IMPLEMENTATION

A. Convolutional Neural Networks

Neural networks were applied using Keras and Tensorflow frameworks for Python 3 programming language. Neural networks used during this research consist of two main parts: convolutional part and the final layers. The convolutional part generalize features and learn data from images. Final layers are tuned for specific classification task. Therefore, originally during ILSVRC neural networks were trained on a large amount of data to recognize thousand of classes. If one wants to use such CNN for classifying other classes of images, one should change final layers corresponding to a new classification task and fine-tune network.

Final layers have been replaced one pooling layer followed by dense layers.

Final neural networks layers presented as:

- Neural Network (InceptionV3, ResNet50, Xception)
- Pooling Layer (Global Average Pooling)
- Dense Layers

Example commands for creation of such networks presented on Fig. 4.

Fig. 3. Scrap metal classes.

```
model = Sequential()
model.add(InceptionV3(include_top=False,
weights='imagenet', classes=classes))
model.add(GlobalAveragePooling2D())
model.add(Dense(units=classes, activation='softmax'))
```

Fig. 4. Example of code snippet for creation a neural network in python.

B. Other machine learning methods

Other machine learning methods use scikit-learn and skimage libraries for Python 3. To use these methods firstly we need to apply local binary patterns. Example of code for extraction features with local binary patterns presented on Fig. 5.

```
lbp = skimage.feature.local_binary_pattern(image, P=8,
R=3).ravel()
histogram = numpy.histogram(lbp, bins=64)
```

Fig. 5. Example of code snippet for extracting features of images in python.

V. THE RESULTS

A. Primary result

Results of applying the methods were described in part II datasets with 9 and 4 classes presented in tables I and II, correspondingly. The authors used following metrics to compare different classifiers classification results: precision, recall, f1 [20].

Having analyzed results presented in table I and II we can conclude that Xception performs the best results between classificators by f1-score in the process of 9 classes classifying, and InceptionV3 performs the best results in the process 4 classes classifying.

TABLE I. CLASSIFICATION OF 9 CLASSES OF SCRAP METAL

Method	Precision	Recall	F1
InceptionV3	0.67	0.67	0.67
ResNet50	0.72	0.73	0.72
Xception	0.77	0.78	0.77
Logistic Regression	0.59	0.60	0.59
Random Forest	0.54	0.52	0.52
Gradient Boosting	0.57	0.58	0.57
Voting	0.59	0.60	0.59

TABLE II. CLASSIFICATION OF 4 CLASSES OF SCRAP METAL

Method	Precision	Recall	F1
InceptionV3	0.88	0.88	0.88
ResNet50	0.87	0.87	0.87
Xception	0.82	0.85	0.83
Logistic Regression	0.65	0.65	0.65
Random Forest	0.67	0.66	0.67
Gradient Boosting	0.70	0.71	0.68
Voting	0.72	0.72	0.72

B. *Upgrated results*

The highest values of the metric f1 in table I and II are not high enough. As stated in previous part, multiple images can be gotten from one railway carriage. All the content of each railway carriage belongs to one class. Thus, the authors can conclude that there are groups of the images taken from same railway carriage are of the same class. Therefore, during testing we can group some images together and be sure that they are taken from same class.

Method of grouping images (Fig. 6) works by this steps:

- Three images taken from each railway carriage. Technically there might be any number of the images, but in this paper minimal odd number was chosen.
- Each of this images runs through classification independently.
- The class commonly-used during the classification results among these three images gets chosen. This class is considered as a result of classification of all three of them.

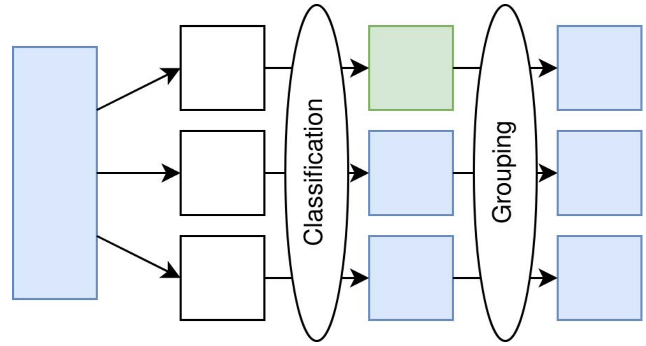

Fig. 6. Visual example of a process of described grouping images for classification method.

In tables III and IV there are presented results for classification of 9 and 4 classes of scrap metal using described grouping method. Having analyzed results presented in table III and IV we can conclude that Xception performs the best results between classificators by f1-score in the process of 9 classes classifying, and both InceptionV3 and ResNet50 perform the best results between classificators by f1-score in the process of 4 classes classifying.

TABLE III. CLASSIFICATION OF 9 CLASSES OF SCRAP METAL WITH GROUPING

Method	Precision	Recall	F1
InceptionV3	0.77	0.78	0.77
ResNet50	0.82	0.82	0.82
Xception	0.88	0.89	0.88
Logistic Regression	0.7	0.72	0.69
Random Forest	0.5	0.5	0.47
Gradient Boosting	0.6	0.56	0.57
Voting	0.58	0.54	0.52

Comparing tables I, II and III, IV it could be noted that f1-score of all methods increases. It allows us to make a conclusion that method used for improving results is effective. Best f1-score were 88 percent on 9 classes and 93 percent on 4 classes, it might be sufficient for automate scrap metal classification depending on the requirements of user.

TABLE IV. CLASSIFICATION OF 4 CLASSES OF METAL WITH GROUPING

Method	Precision	Recall	F1
InceptionV3	0.92	0.93	0.93
ResNet50	0.93	0.93	0.93
Xception	0.88	0.92	0.89
Logistic Regression	0.75	0.77	0.74
Random Forest	0.86	0.86	0.85
Gradient Boosting	0.76	0.8	0.75
Voting	0.76	0.8	0.75

VI. Conclusion

Automated solution for preprocessing images from dataset using Z3 Solver was proposed, which allows getting classification ready images from original photographs. Application of different machine learning methods in task of scrap metal classification was researched. Different classifiers were tested and best ones were determined. Furthermore, method of grouping images was tested and its results have improved in comparison to classifying without using this method. It should be noted that classification results were obtained when dataset didn't have enough number of images for neural networks and machine learning methods training. An increase in the training sample size will provide an increase in the accuracy of classifiers.

References

[1] F. Chollet, Deep Learning with Python, 2017.

[2] A. Kemal, "Exudate Detection for Diabetic Retinopathy with Circular Hough Transformation and Convolutional Neural Networks," Expert Systems with Applications, vol. 114, pp. 289–295, 2018.

[3] E. Cetinic, T. Lipic, and S. Grgic, "Fine-Tuning Convolutional Neural Networks for Fine Art Classification," Expert Systems with Applications, vol. 114, pp. 107–118, 2018.

[4] I. I. Eremenko and I. S. Oliunina, "Use of Machine Learning Methods for Solving Problem of User Identifying by Keyboard Handwriting," Int. Russian Automation Conf., 2019.

[5] I. Bikmullina, N. Andreyanov, and M. Medvedev, "Stand for Development of Tasks of Detection and Recognition of Objects on Image," Int. Russian Automation Conf., 2019.

[6] N.-N. Zhou, A. G. Constantinides, G. Huang, and Z. Shaobai, "Face recognition based on an improved center symmetric local binary pattern," Neural Computing and Applications, vol. 30, pp. 3791–3797 2017.

[7] M. Papakostas and T. Giannakopoulos, "Speech-music discrimination using deep visual feature extractors," Expert Systems with Applications, vol. 114, pp. 334–344, 2018.

[8] M. Robalinho and P. Fernandes, "Software 2.0 for Scrap Metal Classification," 16th Int. Conf. on Informatics in Control, Automation and Robotics, pp. 666–673, 2019.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv:1506.02640 [cs], 2016. [Online]. Available: http://arxiv.org/abs/1506.02640.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," arXiv:1311.2524 [cs], 2014.

[11] Keras Applications. [Online]. Available: https://keras.io/api/applications/.

[12] Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," ArXiv:1512.00567 [cs], 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385 [cs], 2015.

[14] F. Chollet, "Xception: deep learning with depthwise separable convolutions," arXiv:1610.02357 [cs], 2017.

[15] C. M. Bishop, Pattern Recognition and Machine Learning, 2006.

[16] L. Breiman, "Random Forests," Machine Learning, vol. 45(1), pp. 5–32, 2001.

[17] J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," The Annals of Statistics, vol. 29, no. 5, 2001.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa. "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.

[19] Z3Prover. [Online]. Available: https://github.com/Z3Prover/z3/wiki.

[20] Classification metrics. [Online]. Available: https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics.