Online Journal of Applied Knowledge Management

A Publication of the International Institute for Applied Knowledge Management

Volume 10, Issue 1, 2022

Handling of "unknown unknowns" - classification of 3D geometries from CAD open set datasets using Convolutional Neural Networks

Georg Schmidt, LIVINGSOLIDS GmbH, Germany, georg.schmidt@livingsolids.de

Stefan Stüring, LIVINGSOLIDS GmbH, Germany, stefan.stuering@livingsolids.de

Norman Richnow, LIVINGSOLIDS GmbH, Germany, richnow@livingsolids.de

Ingo Siegert, Institute for Information Technology and Communications, Otto von Guericke University Magdeburg, Germany, ingo.siegert@ovgu.de

Abstract

This paper refers to the application of Convolutional Neural Networks (CNNs) for the classification of 3D geometries from Computer-Aided Design (CAD) datasets with a large proportion of unknown unknowns (classes unknown after training). The motivation of the work is the automatic recognition of standard parts in the large CAD-based image data set and thus, reducing the time required for the manual preparation of the data set. The classification is based on a threshold value of the Softmax output layer (first criterion), as well as on three different methods of a second criterion. The three methods for the second criterion are the comparison of metadata relating to the geometries, the comparison of feature vectors from previous dense layers of the CNN with a Spearman correlation, and the distance-based difference between multivariate Gaussian models of these feature vectors using Kullback-Leibler divergence. It is confirmed that all three methods are suitable to solve an open set problem in large 3D datasets (more than 1000 different geometries). Classification and training are image-based using different multi-view representations of the geometries.

Keywords: Open set problem, unknown unknowns, Convolutional Neural Networks, 3D geometries, Metadata, Spearman correlation, Kulback-Leibler divergence.

Introduction

In the task of manual preparation, virtual training procedures for assembly tasks in the automotive industry are created based on the 3D design data (Computer-Aided Design (CAD) data) of the manufacturers (Leu et al., 2013). Unfortunately, the CAD data cannot be processed automatically because it is often faulty, which is mainly caused by the human factor. For example, there are no uniform material assignments in this data, the color information is chosen rather randomly, and the part labels and designations are also not uniform. Furthermore, it is often observed that geometries at the same position are modeled several times which have to be corrected manually. This means that labels or material mappings of the individual geometries in the construction data set were forgotten or incorrectly assigned during the manual design cycle. Especially for small and repetitive geometries, this often means a high workload with highly repetitive work steps (Horejsi, 2015). It is, therefore, reasonable to develop an automatic AI-based algorithm especially

for frequently occurring standard parts. These standard parts are components such as cable clips, screws, or connectors, which are often used by the manufacturer. This means the manufacturers also use the same or similar components in different models, for example, to fasten cables to the car body. The components, therefore, always have the same or similar geometries (Gann, 1996; Klug, 2013). The development of an algorithm that supports the process of standard part identification is the motivation of this work. Therefore, data sets with more than 1000 different geometries (i.e., screws, nuts, sheet metal parts, cables, etc.) are used, which together form the data set of an entire vehicle (Faath & Anderl, 2016). The objective of this work is to automatically extract, classify and label specifically selected geometries with the same or similar shape from the dataset. Classes such as screws, plugs, standard parts, or cable clips (plastic parts for attaching the cable bundle to the car body) are considered here. The background for automating these work steps is the further usability of the data sets. Especially for the subsequent steps in the production of a vehicle, the order, and structure, as well as consistent labeling of the geometries is the basis for further usage (Minow et al., 2020). For the AI-based identification of these standard parts, a supervised learning approach is utilized. Hereby, given examples with known labels are used to train a Convolutional Neural Network (CNN). CNNs are a subtype of neural networks that can generate feature vectors or feature spaces by reducing dimensionality and convoluting a wide variety of input data. These features are then commonly used to distinguish and classify different datasets. Besides speech and text, images are most commonly used as input data for these kinds of classifiers (Chauhan et al., 2018).

The main focus of this paper is on the problem of "Open World Recognition" (Bendale & Boult, 2015b) or "Open Set Recognition" (Bendale & Boult, 2015a; Scheirer et al., 2013). It describes the problem of neural networks being not able to classify data unrelated to any of the pre-trained classes as not belonging to any pre-trained class, which can be paraphrased as "classifying unknowns as unknowns". Neural networks achieve high accuracies in classifying labeled and structured image datasets (Su et al., 2015), but they lack the "ability" to recognize what they do not know (Bendale & Boult, 2015b). While it is possible to train an additional class of known unknowns, it is impossible to train all kinds of inputs of an open world as input. This means that for each neural network that needs to perform a classification task in an open scenario (open set), there will be a set of unknowns that were not included in the training data (Bendale & Boult, 2015b). To prevent the classification of false positives from the open set by the CNN using the Softmax classification layer, three approaches for a respective second criterion are investigated in this work. The remainder of this paper is structured as follows. First, the state of the art in geometry-based classification and recent discussions on the open set problem are referred to. Afterwards, the three proposed methods to overcome the open set problem on geometrical data are introduced. The next section describes in detail the experimental design of the experiments followed by a result section. The paper is then completed with a conclusion and discussion.

State of the Art

Working with 3D data as input data for the training of CNNs is part of current research. In the literature, training methods, based on 3D data, are distinguished from training methods that use a multi-view representation of the geometries. Classification using 3D geometries means accessing the three-dimensional composition of the data directly by using polygon structures, voxels, point

clouds, or geometry metadata (Charles et al., 2017; Feng et al., 2019,). The other approach is to use multi-view representations (Hamdi et al., 2021; Jing et al., 2020; Li et al., 2020; Su et al., 2015). Thereby images of the 3D geometry are taken from different angles and are used as input data for the CNN. Viewing and training with 2D images instead of 3D data sets results to a certain extent in advantages in performance and computation time (Su et al., 2015). The performance advantage is explained by the fact that the computational effort with 2D data is significantly lower than that with a 3D data set. Furthermore, there are investigations on the achieved accuracy of the meshes with the use of 3D data compared to the use of multi-view representations. In (Qi et al., 2016), it was found that 3D-based CNNs performed on average 7.3% worse than CNNs utilizing a multi-view approach. Another advantage of the multi-view approach is that several images of a 3D geometry can be acquired, and thus a network can be trained only with multiple images of a single geometry. Especially with geometries, which appear only once in a data set and have no "similarity" with others, this is a clear advantage. So, regarding the application in this paper this means if the same cable clip is used several times in a car, the goal is to train and automatically recognize it at any position in the vehicle. However, since the geometry only exists once, in the case of training with 3D data, only the same clip is used for training several times, which can quickly lead to overfitting. In a multi-view approach, different images with different viewing angles of a single geometry can be acquired and used for training. This approach avoids overfitting. Due to the advantages of the multi-view approach compared to the use of point clouds or meshes, the training is performed with multi-view image data of the geometries in this paper. As a representation method, a normal mapping of the surface is used. As solutions to the problem of classifying unknown unknowns in an open set recognition task, various approaches are known from the literature (Bendale & Boult, 2015a; Chen et al., 2019; Hassen & Chan, 2020; Uçar et al., 2017). All of the mentioned approaches are based on the assumption that the information, whether a sample belongs to a known or an unknown class, is decoded in the generated features of the CNN, but it cannot be provided by using only a Softmax activation function in the last layer (Hassen & Chan, 2020). Instead, an approach is taken in which the features of the layers before the output layer are used for the computation of a probability distribution (Hassen & Chan, 2020) or also for the classification utilizing another ML algorithm (Uçar et al., 2017). This leads to a second criterion, in which after the classification by the CNN, another decision layer is used to get the final decision whether the current sample is actually a part of a known group or an unknown part of the open set, and thus resulting in a false-positive classification. In this work, this approach is adopted and evaluated for the classification of 3D geometries. The principle of the first criterion remains the same and is based on the image classification of the multi-view representations employing the Softmax output layer of the CNN. For the following second criterion (additional information to the Softmax output layer), three approaches are developed and compared regarding their feasibility in solving the current open set problem.

Experimental Design

2D Image Generation from 3D CAD Data – Knowing the Knowns

For this application, CAD data was used as training data. To preserve the knowledge of this 3D data, specific pre-processing steps were conducted. Firstly, a multi-view approach was used, which has been shown in the literature to be superior compared to point clouds and mesh nets both in

terms of recognition accuracy and computation time (Qi et al., 2016b; Su et al., 2015). For multi-view representations, the 3D CAD geometries are represented as 2D images from different viewpoints and then used as input data for the CNN. This allows a higher resolution to be used as opposed to using 3D data. Furthermore, the multi-view approach prevents overfitting since multiple images (with different information) are created from a single geometry. The following steps are performed to generate the images: At the beginning, a principal axis transformation is performed with each geometry to orient the geometry along the largest axis at the coordinate origin. Then, the smallest possible bounding box is placed around the geometry. The dimensions of this bounding box (height, width, depth) are then used as metadata for later classification. Although, the generation of 3D geometries, the mapping of them as 2D multi-view images, and the calculation of the metadata from the bounding box can be done automatically, the resulting images must be manually labeled (MX10screw, clip, etc.) and care must be taken that there is not a severely underrepresented or overrepresented class. Otherwise, this can lead to under- or overfitting. As a final step in the creation of the "Knowns", the data still needs to be split into training and testing data. In this work, two thirds (2/3) of the data is used for training and one third (1/3) for testing.

Data Generation and Assessing the Knowledge Representation

For the generation of the training images, an internal tool of the company LIVINGSOLIDS GmbH is used. In order to further guarantee optimal preservation of the 3D information also in the 2D data, the surface rendering (i.e., the representation of the surface structure) is one crucial aspect. In this work two common methods are used for this, a gray value image and optical coloring using the normal vector of the surfaces, denoted as surface normals in the following (surface coloring), see Figure 1.

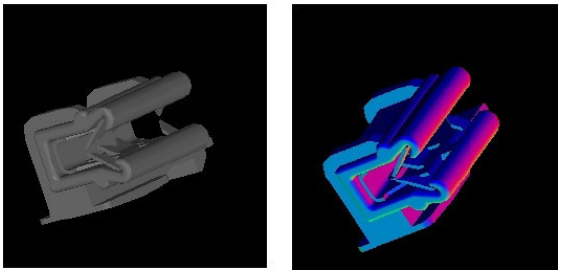

Figure 1. Two Types of 2D Data Generation, Grayscale Image (Left) and Surface Coloring (Right).

As shown in Figure 1, the normal mapping shows a more detailed depth impression. This assumption could be confirmed by the calculation of the entropy of the two types of image generation. The mean value of the entropy for normal mapping is E=3.5, approximately 1.16 times higher than the mean value for the grayscale mapping (E=2.34). This difference in information content is also evident when training an initial CNN (MobileNetV3Large from Google) consisting of 4 blocks with a convolutional layer, a max pooling layer, two dense layers, and a dropout layer each. The difference in recognition accuracy on the test data here is 0.2% after 16 epochs, with 97.2% accuracy from normal mapping vs. 97% with gray values. The difference found by entropy does not show up as strongly in the later classification. However, since the creation of the normal

mapping does not require significant computational effort compared to the gray values, the normal mapping is used for further experiments.

Management of Knowledge by Utilizing Transfer Learning

For the classification architecture, pre-trained models are used, which have been trained using large datasets from the field of object recognition (Russakovsky et al., 2014). A suitable dataset is represented by ImageNet, which consists of approximately 14 million images. Transfer learning is based on the idea of inductive training (Thrun & Pratt, 1998). Here, the first layers of a pre-trained model are used to generate generic features for object description. To adapt to the current dataset, one or more classification layers are then added, consisting of dense layers and a Softmax layer. In the Softmax layer, the number of classes to be recognized is encoded. Afterwards, this classification layer is trained using the current data. In order to select the optimal model for transfer learning with respect to the task at hand, a corresponding training data set was created and the various available pre-trained model architectures were trained over three epochs with the created data set. Afterwards, both the recognition accuracy and the number of operations per run were analyzed. Here, only the added classification layer and the Softmax layer were trained. The feature generation comes exclusively from the architectures pre-trained using Image Net. For the later application, not only a high recognition rate is crucial but also as little computation time as possible, so that the automatic recognition of the different standard parts offers an advantage over manual processing. In total, 26 different models from 8 different architecture families (MobileNet, EfficientNet, DenseNet, Inception, ResNet, NASNetMobile, VGG, and Xception) were compared in this work (Krishna & Kalluri, 2019). The results obtained are shown in Figure 2.

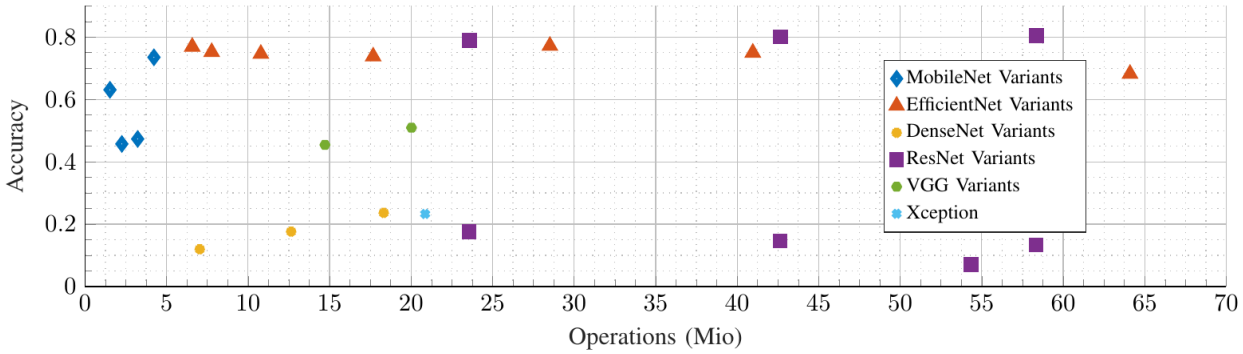

Figure 2. Comparison of the Performance Regarding the Needed Operations for Different Network Architectures on the Utilized Dataset of Cable Clips.

The evaluation of the results shows that model architectures optimized for lightweight mobile applications in particular can still achieve very good results (MobileNetV3Large, 73.5% training accuracy, 4.2 million operations). Larger model architectures (i.e., more operations) achieve only slightly better results here (EfficientNetB6, 74.9%, 4.1Mio operations). Therefore, the MobileNetV3Large model is used for further adaptation in this paper (Howard et al., 2019). Subsequently, a suitable architecture must now be developed that provides reliable results after transfer learning. This means defining layers that need to add to the existing convolutional base of the MobileNetV3Large and that are trained onto our previously defined dataset. For this purpose,

3 models with increasing complexity are experimentally tested. All three approaches are based on the MobileNetV3Large architecture and are shown in Figure 3.

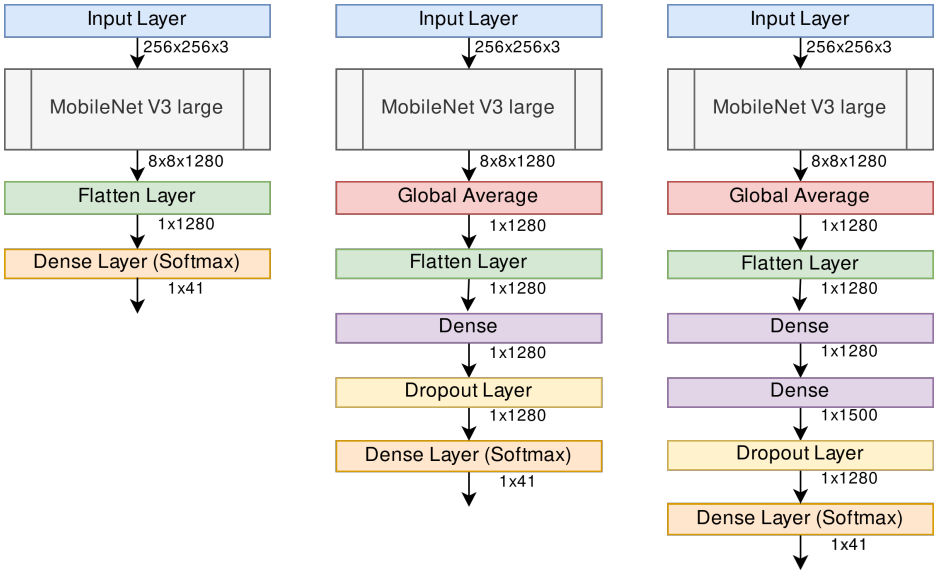

Figure 3. Illustration of the Three Utilized Classification Layer Designs for the Transfer Learning Step.

In subsequent transfer learning, the added classification layers are first trained over ten epochs with a learning rate of 0.0001. Then, fine-tuning is performed over five epochs with 1/10 of the original learning rate. Here, the upper layers of MobilenetV3large are set as trainable and thus also tuned. In terms of recognition accuracy on the training data, all models show a high performance of over 99%. However, with respect to the recognition accuracy on the validation set, there are stronger differences that indicate overfitting, especially in architecture 1 (Figure 3, left). Adding dropout, pooling, and a dense layer can counteract the overfitting in architecture 2 (Figure 3, middle). Adding another dense layer (architecture 3, Figure 3 right) again leads to worse performance. Here it can be assumed that underfitting now takes place. Thus, model architecture 2 with very good performance has been selected for further analyses.

Data Generation

For further evaluation, in total three data sets are considered. The first data set A only contains cables, clips, connectors, and add-on parts of the cable bundles. The data sets B and C contain all geometries of the respective vehicle (i.e., screws, end-plates, steering wheel, cables, engine parts, transmission parts). The cable clips, which are also used for the training in the further process, originate from data set A. They are used to attach cables to the body and have been selected here as examples of standard parts that exist in a large variety but are reused in different vehicles. Data sets B and C are used to verify whether the generalized training approach is suitable for extracting these standard parts from other data sets as well. For the classification of the data sets, an image data set consisting of 30 images each is created from each geometry in the 3D datasets. The

principle is illustrated in Figure 4 using an example data set (Bloodhound Education, 2022) with the generation of images of three selected geometries.

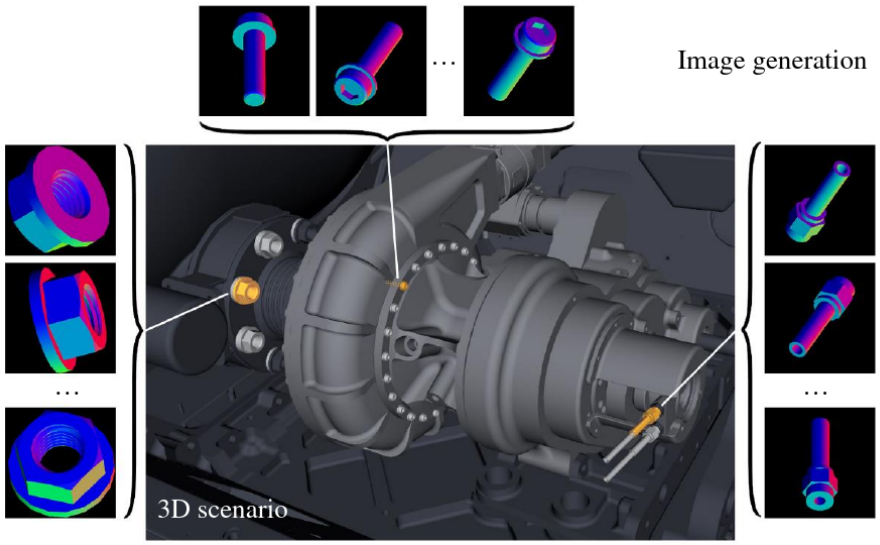

Figure 4. Image Generation From Multiple 3D Geometries

Proposed Methods for Open-Set Recognition

In the context of the application discussed in this paper, the already mentioned topic of unknown unknowns plays a major role. Due to the high similarity of standard parts and special parts, which do not all have their own class but also should not be (falsely) recognized as standard parts. Three different approaches for the recognition of unknown unknowns are developed and evaluated.

Metadata Comparison

The first of the three approaches developed takes advantage of 3D geometries over pure image data. Usually, 3D geometries have metadata information (height, width, depth, surface area), providing additional knowledge for the classification process. Therefore, the first approach is to test whether incorporating this metadata leads to an improvement in the identification of unknown unknowns, i.e., the reduction of false-positive results. The training and subsequent classification steps in this first approach are based on different images of a given geometry. If a geometry sample is recognized by the Softmax output layer of the CNN during testing, the metadata of this geometry is compared with the metadata of the (known) geometry of the corresponding class from the training set using a basic comparison in the absolute values of the height, width, depth and surface specifications. If the metadata match between the two geometries is greater than a threshold of 95%, then the geometry is finally classified as the given class, if not, the classification is rejected.

Feature Vector Correlation

The second approach uses the output of a layer before the final output layer as features. Features from previous layers of the CNN can be generated using Keras's functional API (Gulli & Pal, 2017). Specifically, the three layers before the final output layer are used here, usually denoted as

the feature dense layer. The output of the feature dense layer is a matrix with the dimension 1280 * "the number of input images in the batch process". In the current study, a total of 30 images of each geometry in the data set are used for the prediction and also for generating an original average feature vector from the training data. This set has been experimentally evaluated with consideration of computation time and accuracy. From this matrix, the average can be calculated for each row and thereby the average feature vector for the whole prediction. Thus, for each geometry in the training data set, in addition to the metadata information, the average feature vector, as well as a Spearman correlation between the different vectors, are stored. For the original average feature vector from the training data, the results of the input images are halved for this purpose and then the average feature vectors of these halves are correlated. The Spearman correlation is calculated by the following equation (Meloun & Militky, 2011, p.661):

$$S_{Cor} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)\,(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{N}(y_i - \mu_y)^2}}$$

In this equation, μ is the respective mean value of the vector x (feature vector from images 1-15) or y (feature vector from images 16-30). The classification of a data set is performed as in method one, with the classification of the Softmax layer (first criterion) and then with the second criterion (in this case Spearman correlation). To calculate the Spearman correlation, first an average feature vector from the third layer before the output layer is determined for all images of one geometry in the dataset. This vector is then compared via Spearman correlation to the average feature vector of the geometry classified by the output layer (from train data). If the correlation deviates by more than 10% from the correlation result of the feature vectors of the original training data, the geometry's classification result is rejected. If a deviation of less than 10% is calculated, the classified geometry is accepted (second criterion).

Comparison of Multivariate Gaussian Models/Kullback-Leibler Distance

The generated feature matrix from the second approach can not only be used to generate an average feature vector used directly for a correlation but it can also be used to form a stochastic distribution. At first, the distance between the distribution of the prediction and the corresponding distribution of the original class is calculated. Afterwards, this distance is used to decide whether the images under consideration belong indeed to the same class or are a part of the unknown unknowns. From the literature, this approach is known from the work of Abhijit Bendale (Bendale & Boult, 2015a), in which the author used different Weibull distribution functions as a second criterion for the solvability of an open set classification with image datasets. In this work, it is examined as a third approach whether this is also suitable for the classification of 3D geometries, but with the difference that no Weibull distribution functions are used. Instead, multivariate Gaussian functions are used, because it can be assumed that the mean value (average feature vector) is highly similar among the used images. The mathematical description of a multivariate Gaussian model is given by the mean vector and the covariance matrix (Meloun & Militky, 2011):

$$f_X(x) = \frac{1}{\sqrt{(2\pi)det(S_x)}} exp\left(\frac{-1}{2}(x-\mu)^T {S_x}^{-1}(x-\mu)\right)$$

For each geometry in a training dataset, the corresponding covariance matrix of every class is stored. As the mean vector μ for the description of the multivariate Gaussian model, the already generated average feature vector is used. For the prediction, the multivariate Gaussian distribution of the recognized class (by Softmax layer, first criterion) and the multivariate Gaussian distribution of the considered geometry images (feature vectors) are compared. If the distance between these two models is too large due to a threshold value, the geometry is not classified and vice versa (second criterion). The distance between the two Gaussian models is calculated using the Kullback-Leibler divergence (Duchi, 2018):

$$D_{KL}(N_0, N_1) = \frac{1}{2}\left(tr({S_{x1}}^{-1} S_{x0}) + (\mu_1 - \mu_0)^T {S_{x1}}^{-1}(\mu_1 - \mu_0) - k + ln\left(\frac{\det(S_{x1})}{\det(S_{x2})}\right)\right)$$

The Kullback-Leibler divergence $D_{KL}(N0, N1)$ is calculated using the covariance matrices S_{X1} and S_{X0} and the mean vectors μ_1 and μ_0 of the two distribution functions. As a threshold, a D_{KL} is experimentally determined using distances between N_0 and N_1. The final threshold is then set to 15, for distances >15 the classification result is rejected and for smaller distance values the classification result is accepted.

Results

The CNN utilized in this study is Google's MobileNetV3Large model (Howard et al., 2019). It was pre-trained with the ImageNet dataset (Stanford Vision Lab, 2022) and then adapted for the image representations of the 3D geometries by using transfer learning and optimized with a fine-tuning using Keras (https://keras.io/). The same CNN architecture is used as the basis for classification in all experiments. The difference between the different classification experiments is only in the use of the respective second criterion, i.e., the decision whether the object is part of the known class or part of a yet unknown class. For the final evaluation, the CNN output is tested against a threshold. If a geometry can be assigned with a confidence higher than 95% to a class, the second criterion, i.e., one of the presented methods, is used to decide whether this sample is correctly classified or belongs to the class of unknown unknowns. The results are compared against manually classified and labeled data. Thereby, it is possible to not only compare how many clips were recognized correctly but also how many clips were not recognized. For the comparison of the presented three methods of the second criterion for decision, all results of the classifications of cable clips with the three data sets are listed in Table 1. Additionally, the number of all included cable clips for each dataset is given. The results of the classification without one of the mentioned second criteria are listed to show the extent of the open set problem when no second criterion is used.

Table 1. Comparison and evaluation of the three methods on three data sets

Data set	**Data set A Cable bundle**	**Data set B**	**Data set C**
Geometries in the data set	6.867	9.757	6.452
Clips	932	484	239
Metadata comparison			
Recognized as clip	803	339	159
Correctly classified	803 = 86% of all clips	339 = 70 % of all clips	153 = 64 % of all clips
false-positive	0	0	6
Feature vector correlation			
Recognized as clip	793	364	172
Correctly classified	793 = 85% of all clips	356 = 73% of all clips	151 = 63% of all clips
false-positive	0	8	21
Comparison of multivariate Gaussian models/Kullback-Leibler			
Recognized as clip	829	320	168
Correctly classified	829 = 88% of all clips	320 = 66% of all clips	161 = 67% of all clips
false-positive	0	0	7
Classification without a second criterion			
Recognized as clip	1316	841	423
Correctly classified	724 = 78% of all clips	263 = 54% of all clips	137 = 57% of all clips
false-positive	592	578	286

When comparing the total amount of all cable clips in percent with the amount of correctly classified cable clips, it can be seen that the number of recognized cable clips is the highest for dataset A. This result was assumed since all geometries for the training originate from this data set. However, still not all clips were classified. The fact that only about >80% of all cable clips were recognized by the classification is because of the small number of training examples, which comprises only 71 different clips. A maximum of 88% could be achieved, e.g., by Kullback-Leibler divergence, which leaves room for further optimization by using more training data. The clips that were "not recognized" are geometries that often occur only once in the data set and have not yet been added to the training data. The results of the classifications from datasets B and C show proportions of up to 73% of the total amount of cable clips. It is, thus, confirmed that the trained cable clips are reused in different data sets and can also be extracted from future sets. Although no cable clips in the training data were taken from data sets B or C, the classifier trained on data set A is able to identify cable clips in these datasets, proving that the generalizability of the classifier is successful. Based on the comparison with the results using only the first criterion (i.e., classification without a second criterion in the table), a significant decrease in the ratio of false-positive classifications can be observed for all three methods. In this case, most false-positive

classifications occur when classifying datasets B and C with the feature vector correlation method. Looking closely at these false-positive classifications, it can be seen that the results are due to geometries that have a high geometric similarity to the trained classes. Two examples of false-positive predictions are shown in **Figure 5**.

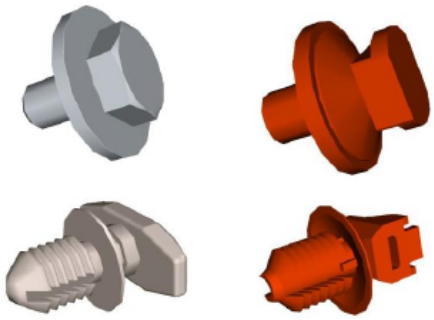

Figure 5. Two Examples of False-Positive Predictions (left) vs. Original Clip (right) in Dataset B

Conclusion and Discussion

The goal of the present work was to develop a framework for the automatic classification of 3D geometries using CNNs that is able to cope with unknown unknowns when being transferred to further datasets. It is confirmed that the application of transfer learning with a pre-trained CNN based on the Image-Net dataset is suitable to subsequently classify multi-view representations of different 3D geometries. Specifically, the convolutional basis of MobileNetV3Large, whose architecture has been extended and re-trained with additional classification layers for the classification of different cable clips was used. The representation of 3D geometries via the rendering of multi-views speeds up the computation effort heavily, since the processing of image data and also the image-based classification with CNNs is more efficient and faster than the processing of each individual geometry several times.

For the application of this architecture in an open set scenario, three different approaches have been implemented as an additional criterion besides the classification by the Softmax output layer of the network. Three approaches were investigated in the current submission. The approaches for the second criterion are the matching of the geometry metadata, a comparison of the feature vectors of a previous classification layer using Spearman correlation, and the distance of the multivariate Gaussian models of these features using the Kullback-Leibler divergence. To analyze and compare the three methods, three different datasets are utilized using the same network architecture. The subsequent comparison of the methods was based on the number of correctly classified clips from the open set datasets and the number of false-positive classifications of unknown unknowns. The evaluation of the three methods for the second criterion proved that all three methods achieved a reduction of up to 100 percent in the number of false-positive classifications from the open set. Slight advantages could be observed when using the first method (comparison of metadata) and the third criterion (distance-based classification with Kullback-Leibler divergence). However, these two methods also encounter proportionally more similarity-based errors than the comparison of metadata. Depending on whether a specific or a general approach for the classification of 3D

geometries from an open set is desired, this aspect should be taken into account. It is possible to confirm that through a one-time effort, namely picking out individual geometries that will be reused in the next datasets, in the future trained CNNs can be used to extract these geometries from new datasets. Especially for manual pre-processing steps to prepare large 3D data sets, this represents a significant time saving and thus an economic advantage. It has to be noted that our approach has some limitations, mainly caused by the application domain. The first criterion (comparison of metadata) is limited to the case that geometric information is available and that the same geometries (in different data sets) have the same dimensions. However, if the size and shape of the metadata remain constant, this criterion is suitable in terms of very reliable and accurate classification rates. The comparison of feature vectors (method 2) and Kullback-Leibler divergence (method 3) on the other hand, is based only on the evaluation of the similarity of the input images, resulting from the generated features and is thus generally applicable.

In terms of knowledge management, the proposed methods aim to avoid misclassification in an open world setting, which is characterized by the fact that the number and variety of not-seen samples is huge in comparison to the seen samples during the classification process. We furthermore discussed on how 3D data can be transferred into the 2D space, to on the one hand reduce the computational effort and on the other hand increase the discriminative power. For the feature representation, we proposed to use multi-view representations. To additionally have a proper knowledge representation, we compared different methods to preserve surface information. For our data, coloring using surface normal turned out to be slightly superior. To further manage previous knowledge in terms of object shapes, we used transfer learning and compared different models and additional layer designs. Finally, three methods on how unknown unknowns can be handled are presented and experiments are conducted. The experiments clearly show that the intervention of human annotators can be reduced, and especially standard parts can be identified automatically. Thus, this work makes benefit of the field of knowledge management for pattern recognition in many aspects and lay out the foundations for further research. In future research, the transferability to further problems, outside the CAD domain, has to be proven. Furthermore, the combination of the proposed solution to identify unknown unknowns with an extended transfer-learning where the identified unknown unknowns are added as new classes to the learning system would be of general interest.

References

Bendale, A. & Boult, T. (2015a). *Towards open set deep networks.* arXiv preprint arXiv:1511.06233.

Bendale, A. & Boult, T. (2015b). Towards open world recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (p. 1893-1902). https://doi.org/10.1109/CVPR.2015.7298799

Bloodhound Education. (2022). *Rocket car.* http://bloodhound1.efar.co.uk/project/car

Chen, W., Wang, Y., Song, J. & Li, Y. (2019). Open set HRRP recognition based on convolutional neural network. *The Journal of Engineering, 2019*(21), 7701-7704. https://doi.org/10.1049/joe.2019.0706

Duchi, J. (2018). Introductory lectures on stochastic optimization. In M. W. Mahoney, J. C. Duchi & A. C. Gilbert (Eds.), *The mathematics of data* (Vol. 25, p. 99-185). IAS/Park City Mathematics Series.

Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

Faath, A, & Anderl, R. (2016) Interdisciplinary and consistent use of a 3D CAD model for CAx education in engineering studies. *Proceedings of the ASME 2016 International Mechanical Engineering Congress and Exposition. https://doi.org/10.1115/IMECE2016-65031*

Feng, Y., Feng, Y., You, H., Zhao, X., & Gao, Y. (2019). Meshnet: Mesh neural network for 3D shape representation. *Proceedings of the AAAI Conference on Artificial Intelligence.*

Gann, D.M. (1996) Construction as a manufacturing process? Similarities and differences between industrialized housing and car production in Japan. *Construction Management and Economics, 14*(5), 437-450. https://doi.org/10.1080/014461996373304

Hamdi, A., Giancola, S., & Ghanem, B. (2021). MVTN: Multi-view transformation network for 3d shape recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (p. 1-11).

Hassen, M., & Chan, P. K. (2020). Learning a neural-network-based representation for open set recognition. *Proceedings of the 2020 Siam international conference on data mining* (pp. 154–162).

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., … Adam, H. (2019). Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (p. 1314-1324). https://doi.org/10.1109/ICCV.2019.00140

Horejsi, P. (2015). Augmented reality system for virtual training of parts assembly. *Procedia Engineering, 100*, 699-706. https://doi.org/10.1016/j.proeng.2015.01.422

Jing, L., Chen, Y., Zhang, L., He, M., & Tian, Y. (2020). Self-supervised feature learning by cross-modality and cross-view correspondences. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* https://doi.org/10.1109/CVPRW53098.2021.00174

Klug, F. (2013) The internal bullwhip effect in car manufacturing. *International Journal of Production Research, 51*(1), 303-322. https://doi.org/10.1080/00207543.2012.677551

Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. *International Journal of Recent Technology and Engineering*, 7(5S4), 427-432

Li, Z., Wang, H., & Li, J. (2020). *Auto-MVCNN: Neural architecture search for multi-view 3D shape recognition.* arXiv preprint arXiv:2012.05493.

Leu, M. C., ElMaraghy, H. A., Nee, A. Y., Ong, S. K., Lanzetta, M., Putz, M., … Bernard, A. (2013). CAD model based virtual assembly simulation, planning and training. C*IRP Annals*, 62(2), 799-822. https://doi.org/10.1016/j.cirp.2013.05.005

Meloun, M., & Militky, J. (2011). *Statistical data analysis: A practical guide.* WPI Publishing.

Minow, A., Stüring, S., & Böckelmann, I. (2020). Mental effort and usability of assistance systems in manual assembly – A comparison of pick-to-light and AR contours through VR simulation. In: Stephanidis, C., Antona, M. (eds) *HCI International 2020 Communications in Computer and Information Science*, (Vol 1224). Springer. https://doi.org/10.1007/978-3-030-50726-8_60

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2016a). Pointnet: *Deep learning on point sets for 3D classification and segmentation.* arXiv preprint arXiv:1612.00593.

Qi, C. R., Su, H., Nießner, M., Dai, A., Yan, M., & Guibas, L. (2016b). Volumetric and multi-view CNNs for object classification on 3D data. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y

Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., & Boult, T. E. (2013). Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(7), 1757-1772. https://doi.org/10.1109/TPAMI.2012.256

Stanford Vision Lab. (2022). *ImageNet Data.* https://www.image-net.org/download.php

Su, H., Maji, S., Kalogerakis, E. & Learned-Miller, E. G. (2015). Multi-view convolutional neural networks for 3D shape recognition. *Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/iccv.2015.114*

Thrun, S. & Pratt, L. (1998). *Learning to learn.* Springer New York. https://doi.org/10.1007/978-1-4615-5529-2

Uçar, A., Demir, Y. & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. *Simulation*, *93*(9), 759–769. https://doi.org/10.1177/0037549717709932

Authors Biographies

Georg Schmidt, M.Sc. Bachelor studies at Otto-von-Guericke-University Magdeburg in Medical Systems Engineering with al thesis on signal processing in ultrasonic devices. Consecutive master studies also in Medical Systems Engineering with a thesis on the application of Convolutional Neural Networks for the classification of 3D CAD data. Currently working as a software developer at LIVINGSOLIDS GmbH Magdeburg. Interests in software development, signal processing, computer graphics, and machine learning.

Stefan Stüring, Dipl. Ing. graduated from Technical University Munich in Mechanical Engineering and started his career as a researcher at Fraunhofer IFF in Magdeburg. He worked in the field of AR & VR to support processes in industrial manufacturing and assembly. In 2005 he spun off the company LIVINGSOLIDS GmbH from the institute and since then leads the industrialization of mixed reality technologies. In recent years AI and neural networks became more important for the handling of mass data from industrial processes which among others led to the work covered by the presented paper.

Norman Richnow, Dipl.Ing. Studies of Computational Visualistics at the Otto-von-Guericke-University Magdeburg. Diploma thesis on "Interactive FEM-Postvisualization as a component of a VR-System: Concept and prototypical realization." Currently working at LIVINGSOLIDS GmbH Magdeburg. Interested in software development, computational geometry, and augmented & virtual reality.

Ingo Siegert, Dr.-Ing. Assistant professor for Mobile Dialog Systems at the Otto von Guericke University Magdeburg. Research interests and publications focus on signal-based analyses and interdisciplinary investigations of human-computer interaction in terms of speaker anonymity, addressee detection, and the utilization of further interaction patterns, such as filled pauses or discourse particles. He has published 100+ peer-reviewed papers on several conferences and various journals and is (co-)organizer of several workshops and conferences.