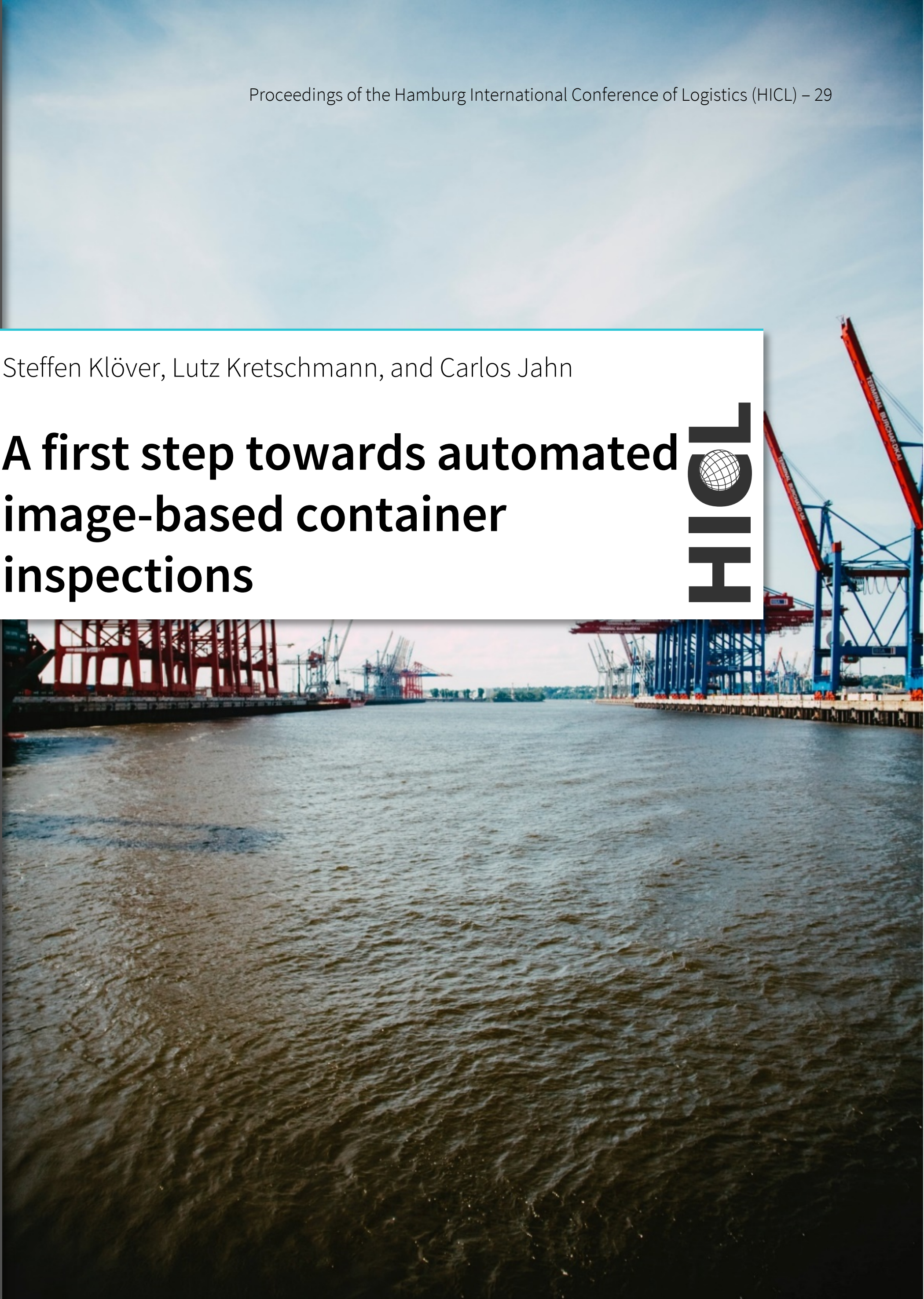
Proceedings of the Hamburg International Conference of Logistics (HICL) – 29

Steffen Klöver, Lutz Kretschmann, and Carlos Jahn

A first step towards automated image-based container inspections



Published in: Data science and innovation in supply chain management
Wolfgang Kersten, Thorsten Blecker and Christian M. Ringle (Eds.)
ISBN: 978-3-753123-46-2 , September 2020, epubli

A first step towards automated image-based container inspections

Steffen Klöver [1], Lutz Kretschmann [1], and Carlos Jahn [2]

1 – Fraunhofer Center for Maritime Logistics and Services CML

2 – Hamburg University of Technology – Institute für Maritime Logistik

Purpose: The visual inspection of freight containers at depots is an essential part of the maintenance and repair process, which ensures that containers are in a suitable condition for loading and safe transport. Currently this process is done manually, which has certain disadvantages and insufficient availability of skilled inspectors can cause delays and poor predictability.
Methodology: This paper addresses the question whether instead computer vision algorithms can be used to automate damage recognition based on digital images. The main idea is to apply state-of-the-art deep learning methods for object recognition on a large dataset of annotated images captured during the inspection process in order to train a computer vision model and evaluate its performance.
Findings: The focus is on a first use case where an algorithm is trained to predict the view of a container shown on a given picture. Results show robust performance for this task.
Originality: The originality of this work arises from the fact that computer vision for damage recognition has not been attempted on a similar dataset of images captured in the context of freight container inspections.

First received: 11. Mar 2020 Revised: 19. Jun 2020 Accepted: 12. Aug 2020

1 Introduction

The availability of empty cargo containers in an appropriate condition is a prerequisite for consignors to load cargo and shipping companies to move containers to their destination subsequently (Olivo, 2005). In order to ensure sufficient availability of empty containers and thus meet regional demand, shipping companies go to great length. Imbalanced trades require repositioning and temporary storage of empty containers in regional depots (Schlingmeier, 2016). Besides storage these empty container depots also carry out essential inspection, maintenance and repair tasks, which ensure that containers are in a suitable condition for loading and safe transport (Port of Hamburg Magazine, 2020).

Upon arrival at a depot, empty containers are subject to a visual inspection process at the gate. The main objective of this process is to separate intact containers, which can go directly into the storage area, from damaged units, which require maintenance and repair. Where a damage is detected the process will further determine different attributes that characterize the condition (including damage type, extent and location). This information is used to arrange appropriate maintenance and repair measures subsequently and also support commercial transactions with the container owner (customer). Thus, the visual inspection of freight containers at depots is an essential first step of the overall maintenance and repair process. Comparable visual inspection processes can also be found at other container terminals, which primarily fulfill transshipment and handling functions within the supply chain.

Today, inspection processes are carried out by experienced staff, which identify damages visually, document them by taking pictures und provide repair proposals. This process set up has several disadvantages. First of all, damage assessments are subjective and can differ between inspectors based on, amongst others, their individual experience. Further, highly skilled inspectors are in high demand making them expensive respectively limiting their availability. Lastly, inspection of containers upon arrival, identifying individual damages and documenting them manually by taking pictures and entering damage characteristics, e.g. in a handheld device, is both error prone and time-consuming.

With significant advances in the field of computer vision over the recent years, automating visual inspection tasks across industries and applications has become principally feasible (Brownlee, 2019). Today deep learning dominates most computer vision applications and provides state-of-the-art performance (Russakovsky, et al., 2015 and Chollet, 2017). This paper is dedicated to applying computer vision algorithms in the context of visual container inspections with the goal to overcome the previously mentioned disadvantages of the current process. It introduces a computer vision model, which uses deep learning methods, to automate the inspection process. Based on a large dataset of images taken during freight container inspections, a deep learning computer vision model is trained and evaluated in its performance.

In order to confirm the applicability of computer vision for an automation of container inspections the overall research process consist of several steps that need to be investigated. Individual steps include identifying containers with / without damages and classifying these by damage type as well as establishing damage location on the container and specifying the affected container component. Within this overall context, this paper focusses on predicting the view of a container (e.g. top, front, or side view) shown on a given image, which is required to identify and classify a damage as well as establishing its location on the container. Future work will focus on computer vision models, which address other tasks in order to complete further steps of the overall research process. Results presented here, in terms of metrics like accuracy and precision, will also serve as a benchmark for assessing the performance of computer vision models for all individual parts of an automated image-based inspection of freight container condition.

The remainder of this paper is structured as follows. Section 2 gives an overview of the inspection process at empty container depots today and describes which subtask this paper focusses on by describing the results of automating the subtask through a deep learning computer vision model. Subsequently Section 3 introduces current deep learning methods used in the context of computer vision and gives a brief overview of related work. The use cases "predict the container view" is covered in Section 4 with individual subsections describing the data used for model training, the chosen model architecture as well as achieved results. Concluding remarks and possibilities for future research are given in Section 5.

2 Container Inspection Process

At the empty container depot arriving containers are subject to a visual inspection to ensure structural integrity and fitness for safe transport. Reliable identification, evaluation and subsequent repair of any damage to the container is a crucial aspect to comply with the Convention for Safe Containers (CSC, 1972). In this context, the industry has reached an agreement to comply with several generally accepted standards:

- Criteria for assessing damage (UCIRC, RCIRC and IICL6)
- CEDEX coding for damage documentation

The CEDEX coding was developed as "1985-87 ISO TC104" and first published in 1989 as ISO 9897 (ISO 9897). CEDEX is used worldwide as a standard way to document damages to freight containers. It consists of four elements to indicate

- "Location" of a damage (4 digits)
- "Component" affected by a damage (3 digits)
- "Damage" type present (2 digits)
- "Repair" measure suggested (2 digits)

Below, Table 1 gives an exemplary CEDEX code and a decoded description of the affected location and component, the present damage and the associated suggested repair measure.

In practice today, the process of damage identification is characterized by a sequence of manual activities, which are merely supported by digital means. In this context, a mix of paper-based documentation steps and the use of hand-held devices result in a high risk of errors and process steps are time-consuming. On some depots the process is at least partially digitalized. Here an inspector assesses the container's condition and documents his findings on a hand-held mobile device supplemented by taking pictures of the container in its entirety and the a closeup of the respective damage. The process requires several standardized pictures to be taken, which cover the outer container walls from different angles as well as the inside (see Figure 3). In case a damage is detected, it is further documented by a close-up detail picture. The data structure used to document information collected during the inspection process follows the CEDEX code. Figure 1 shows an illustration of the inspection process.

Table 1: Example of CEDEX code and decoded description

	CEDEX example	Description
Location	DH4N	D → Door end (rear) H → Higher portion (upper) 4 → Right-hand side corner post N → element not used in this case
Component	CFG	Fittings located at the corners of containers providing means of supporting, stacking, handling, and securing the container
Damage Type	CU	Component is damaged by being cut
Repair	RP	Remove and replace the complete cross-sectional profile of a component over its entire length and width

The current inspection process - as described before – and subsequent maintenance and repair processes set the framework for an automation of container inspections by using computer vision models, as covered in this paper. Accordingly, automation of the entire visual inspection process or individual parts would require the same output as today to be compatible with the current processes on the depot. This leads to the main research question: is it possible to build a deep learning model that is capable to distinguish damaged from undamaged containers based on a given image and further, in case a damage is identified, determine (or rather predict) the individual CEDEX elements for this damage.

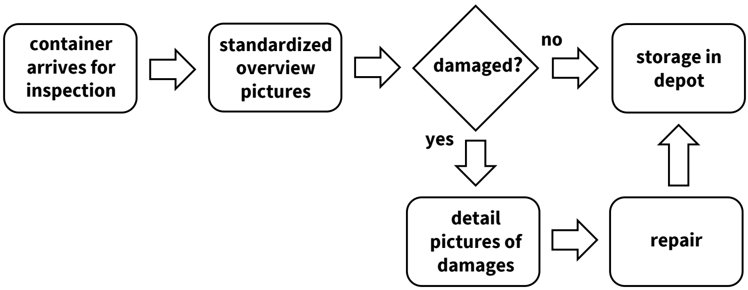


Figure 1: Container inspection process

Due to the different nature of individual sub-problems, a combination of models will most likely be required to meet the overall requirements. Separate models would be dedicated to solve different subtasks. In that sense one model might distinguish damaged containers from intact units while other models are used to predict damage location, damaged component, damage type as well as the appropriate repair measure. Within the overall scope of the described principal research question, this paper addresses a first subtask: automatically predicting the first of four characters that constitute the CEDEX element "damage location". This character specifies the respective face of the container with eight main distinctions. Thus, based on an image taken by an inspector during the inspection process on the terminal, the model has to predict which view of a container is shown.

3 Deep Learning for Computer Vision

Artificial intelligence and in particular machine learning have seen a significant increase in use over the past few years. Machine learning comprises a number of different algorithms that have been developed to learn correlations through pattern recognition in data sets and use these correlations to make predictions for new, previously unknown data (Nelli, 2018). Today, most successful applications of machine learning are from the sub-field of supervised learning. Here a large number of labeled examples (combination of input data and desired output) are processed automatically in order to learn statistic correlations that characterize the relation between input and output. Subsequently, these relationships can be used in the sense of decision rules when predicting the corresponding output for a given input (Müller and Guido, 2017). More recently, deep learning has emerged as a new subfield in machine learning, with performance of deep learning models significantly exceeding the performance of classical machine learning algorithms in various supervised learning problems. Facilitated by ever-increasing computing power and data volumes, deep learning has thus enabled remarkable breakthroughs in applications that process e.g. image, text or sound data (Le Cun and Bengio 1995, Goodfellow, et al., 2016 and Le Cun, et al., 2015).

Computer vision is a subfield of artificial intelligence concerned with the automated extraction of information from visual image data. From inferring the depth of a scene in stereo vision to deducing the presence of an object in recognition, computer vision is a multidisciplinary field with various approaches to address different kind of problems (Prince, 2012). Today it is powering applications like image search, robot navigation, face detection, autonomous driving and many more (Szeliski, 2010). Achieving the performance of human perceptual vision has been a challenge for decades (Prince, 2012). Through the complexity of visual data, objects can appear in any pose and are often partially occluded by other objects. Recognizing different objects in images requires the extraction of visual features which provide a semantic and robust representation. Due to diversity in appearance, illumination condition and backgrounds it is difficult to manually design feature descriptors like SIFT (Lowe, 2004) and HOG (Dalal and Triggs, 2005) – examples of pre-deep learning computer vision methods - to perfectly describe all kinds of objects (Zhao, 2018). Recognizing objects in images can be treated as a pattern recognition problem, where new images are interpreted based on prior images in which the context is known. This process can be divided into two parts. In learning, the relationship between image and content is modeled. In inference, this relationship is exploited to predict the content of new images (Prince, 2012).

Computer vision methods for image analysis in principal represent a suitable solution for automated detection of damaged containers – the field of investigation in this paper. Under this approach a deep learning model is trained to predict the correct classification of e.g. damage type or damage location in each case based on a dataset of labelled examples - a combination of digital image and the corresponding label. The learned relationship is then used to predict the correct label for a new image taken at a container depot. This way automating the inspection process through computer vision can be framed as an object recognition task. Recognizing objects in images requires the extraction of features that provide a semantic and robust representation. For image recognition problems of this kind, convolutional neural networks have proven a superior performance compared to "traditional" computer vision methods in terms of error rates (Krizhevsky, et al, 2012). Convolutional neural networks, a special type of deep artificial neural network, are based on a hierarchical multi-stage structure, capable of learning multilevel representations from pixel to high-level sematic features. Compared with traditional "shallow" machine learning, this deeper architecture provides an increased expressive capability (Zhao, 2018). Convolutional neural networks have demonstrated superior performance on a variety of object recognition tasks, like image classification and object detection (Russakovsky, et al., 2015). An important contribution towards the performance of latest generation models has come from the ImageNet Large Scale Visual Recognition Challenge. This annual computer vision competition for both object detection and image classification tasks had a large impact on the evolution of computer vision during the 2010s with the achieved success primarily driven by very large and deep architectures of convolutional neural networks in combination with the computing power provided by graphical processing unit hardware (Brownlee, 2019). As a consequence of this development deep neural networks nowadays perform better than or on par with humans on image classification tasks, especially if good quality image data is available (e.g. limited distortion, noise, blur) (Dodge and Karam, 2017).

The application of computer vision methods has been the subject of scientific publications in several different fields up to this day. A selection of research comparable to this paper regarding the methodology and use cases is discussed hereafter. Jaccard, et al. (2016) apply computer vision methods in the context of X-ray cargo inspections. These inspections take place at border crossings to interdict trade in illicit or security related goods such as contraband, drugs or weapons. In view of rising cargo volumes, increasing regulations and the need for more efficient handling processes at land borders and in ports, automation of parts of the inspection workflow is proposed as a suitable solution. If implemented successfully it could reduce processing times and enable expert operators to focus only on high-risk containers and suspicious images. The authors propose a framework for automated cargo inspection consisting of several modules and use a large dataset of X-ray cargo images for training and evaluation. Besides classical machine learning methods, such as a random forest algorithm used in a car detection module, the setup also includes state-of-the-art deep learning approaches. In particular the authors evaluated convolutional neural network architectures with different depth (number of layers), down-sampling of pixels (to make training computationally tractable), and 3X3 filters trained on 12,000 images from a negative class (no threat) and 12,000 positive class images with a synthetic threat build into the picture. Compared to other computer vision methods the authors where able to demonstrate significantly better detection performance with the suggested deep learning set up, achieving a false positive rate of 0.80% for a detection rate of 90%.

Another recent example can be found in the work done by Patil, et al. (2017) who employed deep learning methods for car damage classification on a relatively small dataset of manually annotated images from the internet. The images show damaged and undamaged sections of cars, enclosing components from a variety of different angles and distances partly similar to container inspection photos. In a multiclass classification with eight classes, including seven damage types and one class for no damage, best results were obtained using state-of-the-art convolutional neural networks pre-trained on the ImageNet dataset as a feature extractor and a fully connected network with softmax activation function as a classifier. With 88.24% accuracy ResNet (He, 2015) performed best. Building an ensemble classifier on top of multiple different pre-trained classifiers was able to boost accuracy slightly up to 89.53%. The work also demonstrates the ability of this approach to localize damage instances at the pixel level. This is achieved by cropping a region of size 100 x 100 around pixels, resize it to 224 x 224 pixels and predicting the class probabilities. Pixels with class probabilities above a threshold can be highlighted respectively.

Other scholars including e.g. Maeda, et al. (2018) and Shihavuddin, et al. (2019) and Perez, et al. (2019) have previously applied state-of-the-art deep learning technologies in different engineering domains in order to recognize structural damage from images, which is somewhat comparable to the use case covered in this paper. Maeda, et al. (2018) focus on damage detection on road surfaces using image data and deep neural networks. They collected and prepared a large road damage dataset and successfully trained a convolutional neural networks model to predict eight different damage types with high accuracy. Another engineering application can be found in Shihavuddin, et al. (2019) who work on detecting surface damages on wind turbine blades. A deep learning-based automated damage detection system for analysis of drone inspection images is proposed with a particular emphasis how advanced data augmentation can increase precision for small training sets. Lastly, Perez, et al. (2019) focus on the use case of automatic detection of building defects from visual images. With an increasing interest in quick and effective means to survey the condition of buildings, developing ways to accomplish this task by computer vision models is presented as a promising alternative. The authors propose a suitable model for their use case based on a pre-trained convolutional neural network classifier and achieve robust results for the task of detecting and localizing building defects.

4 Use Case "Predicting Container View"

Automatic image recognition and evaluation has the potential to increase the efficiency of processes in seaports and terminals. For this reason, terminal operators have been using OCR (Optical Character Recognition) systems for many years to automatically identify incoming truck's license plates as well as container numbers. (HHLA n.d.) In comparison, automatic detection of damaged containers by computer vision methods goes beyond state of the art.

In this section a computer vision module is introduced, which predicts the first character of the CEDEX damage location code based on a given image. The problem is framed as a multi-class classification problem and state of the art deep learning computer vision models are implemented and evaluated. The reviewed literature demonstrated that supervised learning approaches utilizing convolutional neural networks show state-of-the-art performance in image-based damage classification even on small datasets. Since there is a large number of annotated images available for the use case of this paper pursuing this approach should provide promising results. In the chosen approach a convolutional neural network is trained based the first character of the CEDEX damage location code provided by the inspector's annotation. Besides showing the general feasibility of the selected approach, a main focus was on investigating the impact of image resolution on model performance and thus to get an estimate of how much information loss through compression is still tolerable. This is of particular importance in the considered context as the available data base consists of high-resolution images. However, training a deep learning computer vision

model will require a certain reduction of image size in order to be computable on today's hardware in an adequate time.
The remainder of this section will firstly introduce the used data set of container images and then describe the implemented deep learning architecture. Lastly results of training the model with different image resolutions and with/without fine-tuning is presented.

4.1 Data

The image data available to this research was taken on an empty container depot over the period of 01-2017 to 03-2020. Individual images were captured as part of the inspection process by experienced staff using a mobile handheld device. Overall the dataset contains 568,120 standard type images of containers taken during inspection or at the depot gate. Each image is provided with a label, identifying the respective container view shown on the picture. Overall the following eight classes are distinguished:

- Front view
- Left view
- Door/rear view
- Right view
- Underside view
- Bottom/floor view
- Insight view
- Roof/top view

Five exemplary pictures for each class are shown in Figure 3. These pictures also provide a first impression of possible challenges to a correct automatic classification by a deep learning model, e.g. due to different perspectives and lighting conditions.

The distribution of class labels in the used data set shows a certain imbalance, which can be another challenge for class prediction with high accuracy. The class door/rear is the majority class representing more than 20% of all images while the label "bottom/floor" is the minority class with less than 5% of all images (see Figure 2). The distributions of class labels found in the overall data set was preserved in all splits used for model training and evaluation (training, validation and evaluation data sets).

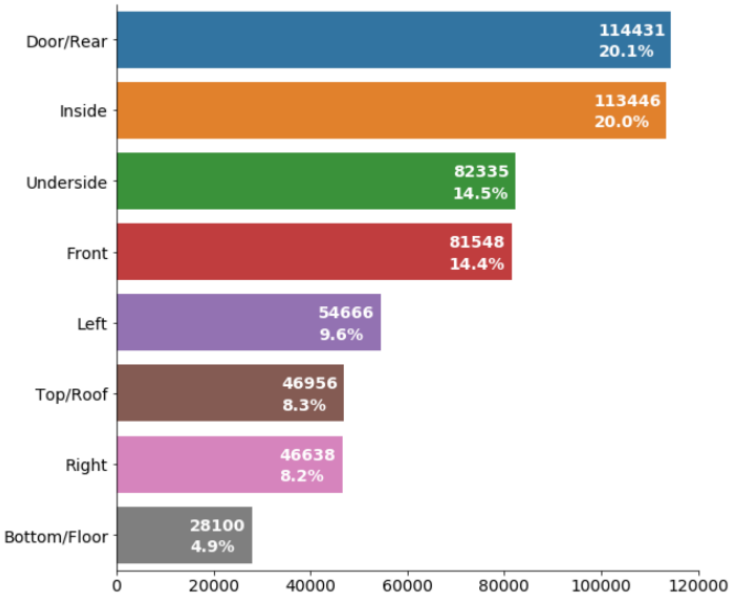


Figure 2: Distribution of class labels in the overall data set

The resolution of the images in the data set varies between 1024 x 768 pixels and 768 x 1024 pixels in height and width. The distributions of image resolution found in the overall data set was also preserved in all splits used for model training and evaluation. Independent of original size (1024x769 or 768 x 1024), images were compressed obtaining equally spaced dimensions in height and width. Further, as part of data pre-processing, pixels were scaled from an initial 0-255 RGB value per color channel to a value between -1 and 1 sample-wise (not averaged over batches).
In the classification experiments 70 % of images in the dataset were used for training and 10 % for validation during training. The remaining 20 % were used to evaluate and compare the performance of trained models.

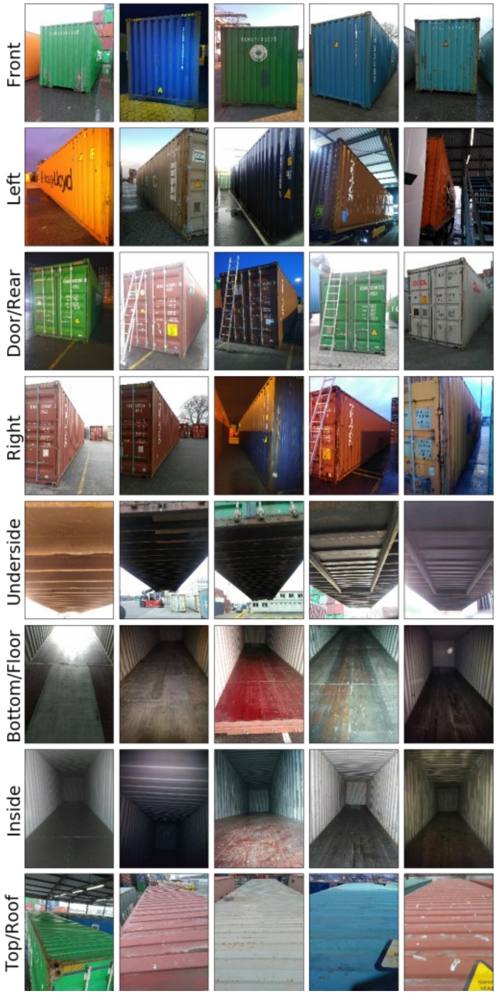


Figure 3: Examples of pictures from the data base representing all classes considered in the use case. Source: HCCR, 2020

4.2 Architecture

The deep learning computer vision architecture used to train a model for the use case “prediction container view” contains of an input layer followed by two main elements, the feature extractor and the classifier.

For the feature extractor the MobileNetV2 architecture (Sandler, 2019) was adopted with weights pre-trained on the ImageNet (Russakovsky, 2015). MobileNetV2 provides a very simple architecture that is specifically tailored for mobile and resource constrained environments by significantly decreasing the number of operations and memory needed while retaining the same accuracy. These properties would principally allow an integration of the computer vision module into mobile devices, which is used during the container inspection process instead of computing on a remote server.

For the classifier a common set up was used, which consists of a fully connected and an output layer. These receive the flattened output of the feature extractor and further match this one-dimensional representation with eight classes in the output layer.

A number of different experiments were considered. First of all, a main focus of this research was on investigating the impact of image resolution on model performance. Accordingly, a high, medium and low resolution set up with pictures resized to 224x224, 112x112, and 56x56 was considered (in effect halving input dims twice). Beyond that distinct experiments were carried out without respectively with fine-tuning the model. The latter included an optimization of the feature extractor weights. Training in all cases started with randomly initialized weights in the classifier (dense / fully connected network with softmax activation layer).

4.3 Results

The performance of the considered classification task “predicting container view” were evaluated by using the metrics accuracy, precision, recall and F1-Score. After each epoch accuracy was measured on the validation set. Training was carried out until performance in terms of accuracy declined on the validation set. At that point all metrics were calculated on the test set.

For each class c_i the number of correctly identified images (True Positives tp_i), and the number of images that were either incorrectly classified as class c_i (False Positives fp_i) or incorrectly assigned to another class (False Negatives fn_i), can be counted individually. The overall Accuracy (1) is defined as the total number of correctly identified image labels across all categories (True Positives tp_i) divided by the total number of images in each class n_i.

$$\text{Accuracy} = \frac{\sum_{\mathrm{i}}^{\mathrm{c}} \mathrm{tp_i}}{\sum_{\mathrm{i}}^{\mathrm{c}} \mathrm{n_i}} \tag{1}$$

Additionally, precision (2), recall (3) and the F1-Score (4) measures were calculated for each class individually and then averaged into a single value (unweighted mean). Calculating the average like this, called macro-averaging, better accounts for class imbalances, since the scores are more influenced by the performance on minority classes (Yang, 1999)

$$\text{Precision} = \frac{1}{\mathrm{C}} \sum_{i}^{\mathrm{C}} \frac{\mathrm{tp_i}}{\mathrm{tp_i} + \mathrm{fp_i}} \tag{2}$$

$$\text{Recall} = \frac{1}{\mathrm{C}} \sum_{i}^{\mathrm{C}} \frac{\mathrm{tp_i}}{\mathrm{tp_i} + \mathrm{fn_i}} \tag{3}$$

$$\text{F1-Score} = \frac{1}{\mathrm{C}} \sum_{i}^{\mathrm{C}} 2 \frac{\mathrm{Precision_i}\ \mathrm{Recall_i}}{\mathrm{Precision_i} + \mathrm{Recall_i}} \tag{4}$$

Table 2 shows the performance of the model on the test dataset for different image resolutions and with/without fine-tuning. Evaluation metrics show that the model performs much worse without fine-tuning layers of the feature extractor. Tracking metrics during training also revealed that the model is not able to generalize well, showing strong overfitting independent of image resolution. The best model accuracy is 0.9753 for 224 pixels for height and width with fine tuning, although the model did not perform much worse using only 112 pixels.

Table 2: Performance of deep learning architectures for predicting the container view

Resolution	Fine-tuning	Accuracy	Precision	Recall	F1-Score
56x56	No	0.7234	0.7411	0.6377	0.6413
112x112	No	0.8145	0.8574	0.7634	0.7828
224x224	No	0.7533	0.9086	0.6366	0.6263
56x56	Yes	0.9404	0.9277	0.9271	0.9252
112x112	Yes	0.9682	0.9594	0.9604	0.9597
224x224	Yes	0.9753	0.9691	0.9674	0.9681

Figure 4 contains the confusion matrix of evaluating the architecture with 224x224 pixels and fine-tuning of the model. A first finding is that class-wise predictions show False-Positives in almost all classes. A possible explanation to this surprising result is wrongly labeled images. Non the less, this aspect warrants a deeper analysis in subsequent research and in particular should the computer vision model be implemented in practice in order to avoid problematic misclassification. Beyond that, there is a number of classes where comparatively many misclassifications occur. This mainly involves:

- Left view misclassified as Front and Right view
- Right view misclassified as Left view
- Front view misclassified as Left view
- Inside view misclassified as Bottom/Floor view
- Bottom/Floor view misclassified as Inside view
- Door/Rear view misclassified as Right view

A visual check of a sample of misclassifications images lead to the following possible explanations. A right / left misclassification can occur if the door or frontside is not visible on the image and thus there is no way of telling if the right or left container side is shown. With respect to distinguishing bottom/floor view from inside view the angle from which a photo is taken is decisive. However, there is no clear separation between this angle in photos of the container floor respectively the container insight. Accordingly, photos with an angle in this gray area are prone for misclassification. A similar effect occurs for a distinction between door/rear and right respectively front and left. One some side views the front/rear is visible but not on all. The same is true for some front/rear views where part of the side is visible on some but not on all. This results in a transition area which is difficult to discriminate for the model.

The applied method shows promising results for inferring the face or view of container which can be used as first step in predicting the location of damages. The results also demonstrate that there are transition areas between the considered categories which can lead to misclassifications. This is a challenging problem which will likely to be intensified in predicting the location of damages in a more fine-grained manner.

On the basis of the results obtained, a number of interesting findings give rise to possible research in the future. First of all, a closer look at classifications errors is warranted. In particular a number of possible causal factors could be investigated:

- Impact of lighting conditions which could result in vanishing boundaries between containers and background.
- Impact of images showing damaged / undamaged containers, which could result in e.g. pictures taken from different angles
- Impact of resolution (1024 x 768 or 768 x 1024) which could lead to different distortions of the object (container) in images

Another promising area of further investigations concerns the use of image augmentation. This includes, amongst other, artificially varying lighting conditions or a padding of images to create original images with a size of 1024 x 1024 by adding a black padding for example.

Predicted \ Actual	Bottom/Floor	Door/Rear	Front	Inside	Left	Right	Top/Roof	Underside
Bottom/Floor	5305	12	15	123	9	10	10	4
Door/Rear	1	22734	16	5	12	78	4	4
Front	12	36	15697	9	292	24	4	8
Inside	215	6	3	22511	2	5	39	5
Left	56	24	543	10	10183	552	27	6
Right	9	45	9	1	414	8650	3	3
Top/Roof	10	5	8	29	20	8	9304	8
Underside	12	24	19	1	1	1	0	16429

Figure 4: Confusion matrix of deep learning architecture (224x224 pixels with fine tuning)

5 Conclusion

The inspection of containers for damages is an essential part of the maintenance and repair process. Since insufficient availability of skilled inspectors can cause delays and result in poor predictability, automation promises the potential for optimization. Automatically recognizing damages through computer vision would speed up the process and enable inspectors to focus their attention on damages that are likely to be anomalous or not sufficiently visible at the surface.

This paper has demonstrated that state-of-the-art deep convolutional neural networks are capable to predict the correct container view shown in an image captured during the inspection process. This corresponds to a first subproblem of predicting damages according to the CEDEX coding for damage documentation and thus automating the overall inspection of containers for damages. Once further research steps are completed it will be interesting to look for other possible fields of application for the achieved results.

Despite achieving a first step in automatizing container damage inspections, it is clear that much remains to be done. Subsequent research will focus on computer vision models that automatically predict the remaining CEDEX code elements. Accordingly, next steps will involve image-based differentiating between damaged and intact containers, the determination of the exact type of damage as well as the affected component of the container. Moreover, since there are various container types with specific damages occurring more or less often, predicting the correct classification for rare cases will be a particular challenge that needs to be addressed.

Acknowledgements

The authors acknowledge the financial support by the Federal Ministry of Transport and Digital Infrastructure of Germany in the framework of the project COOKIE (project number 19H19006B).

References

Brownlee, J., 2019. Deep Learning for Computer Vision - Image Classification, Object Detection and Face Recognition in Python

Chollet, F., 2017. Deep Learning with Python, Manning Publications Co.

CSC, 1972. The International Convention for Safe Containers

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection

Dodge, Samuel and Karam, Lina 2017. A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions. 26th International Conference on Computer Communications and Networks, ICCCN

European Environment Agency (EEA), 2014. Annual European Union greenhouse gas inventory 1990-2012 and inventory report 2014. [online] European Environment Agency (EEA). Available at: <http://www.eea.europa.eu/publications/european-union-greenhouse-gas-inventory-2014/greenhouse-gas-inventory-2014-full.pdf> [Accessed 18 June 2014].

Flint, D.J., Larsson, E., Gammelgaard, B. and Mentzer, J.T., 2005. Logistics Innovation: A customer value-oriented social process. Journal of business logistics, 26(1), pp.113–147.

Goodfellow, I., Bengio, Y., and Courville, A., 2016. Deep Learning, MIT Press, url http://www.deeplearningbook.org

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition

HHLA n.d. Geschichte des HHLA CT Burchardkai. Online: https://hhla.de/de/container/burchardkai-ctb/geschichte-ctb.html. Last accessed 29.05.2020

IICL 6 Inspection Criteria (IICL6) Sixth Edition (IICL-6) - Institute of International Container Lessors

ISO 9897 :1997 Freight containers — Container equipment data exchange (CEDEX)

Jaccard, N., Rogers, T.W., Morton, E.J., Griffin, L.D., 2016. Tackling the X-ray cargo inspection challenge using machine learning. In: Ashok, A and Neifeld, MA and Gehm, ME, (eds.) Anomaly Detection and Imaging with X-Rays (ADIX), 98470N (May 12, 2016); doi:10.1117/12.2222765. (pp. 98470N). Society of Photo-Optical Instrumentation Engineers (SPIE): Bellingham, Washington, USA.

Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. Imagenet classifcation with deep convolutional neural networks. Adv. in Neural Information Processing Syst. 25.

Le Cun, Y., Bengio, Y., 1995. Convolutional networks for images, speech, and time series, in The Handbook of Brain Theory and Neural Networks, M. A. Arbib, Ed. Cambridge, MA: MIT Press

Le Cun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. Nature 521, 436–444 (2015). https://doi.org/10.1038/nature14539

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., Omata, H., 2018. Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone, Computer Aided Civil and Infrastructure Engineering

Müller C.A., Guido, S. 2017. Introduction to Machine Learning with Python. O'Reilly Beijing Boston Farnham

Nelli, F., 2018. Python Data Analytics. Apress Media, California

Olivio, A., Zuddas, P., Di Francesco, M., Manca, A., 2005. An Operational Model for Empty Container Management

Perez, H., Tah, J.H.M., Mosavi, A., 2019. Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. Sensors 2019, 19, 3556.

Prince, S., 2012. Computer Vision: Models, Learning and Inference. Cambridge University

Port of Hamburg Magazine, 2020. A Cookie for Containers, Port of Hamburg Magzine, Page 14

Plowman, E.G., 1964. Lectures on elements of business logistics. Stanford University, Graduate School of Business.

Refrigerated Container Inspection and Repair Criteria (RCIRC) 3rd Edition 2005. Institute of International Container Lessors

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., Frei-Frei, L., 2015. ImageNet Large Scale Visual Recognition Challenge

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L., 2019. MobileNetV2 Inverted Residuals and Linear Bottlenecks

Schlingmeier, J., 2016. Koorperation in der martimen Leercontainerlogistik, Fraunhofer Verlag

Shihavuddin, A.S.M., Chen, X., Fedorov, V., Christensen, A.N., Riis, N.A.B., Branner, K., Dahl, A.B., Paulsen, R.R., 2019. Wind Turbine Surface Damage Detection by Deep Learning Aided Drone Inspection Analysis. Energies 2019, 12(4), 676

Szeliski, R., 2010. Computer Vision: Algorithms and Applications, Springer Verlag

Unified Container Inspection & Repair Criteria (UCIRC), Revision 2, 2000. International Chamber of Shipping

Yang, Y., Lui, X., 1999. A re-examination of text categorization methods

Zhao, Z., Zheng, P., Xu, S., Wu, X., 2018. Object Detection with Deep Learning: A Review