Towards Graph Representation based Re-Identification of Chipwood Pallet Blocks

Simon Klüttermann[1*], Jérôme Rutinowski[2*], Christopher Reining[2], Moritz Roidl[2], and Emmanuel Müller[1]
[1]TU Dortmund University, Germany – Chair of Data Science and Data Engineering
[2]TU Dortmund University, Germany – Chair of Material Handling and Warehousing
Emails: {simon.kluettermann; jerome.rutinowski; christopher.reining; moritz.roidl; emmanuel.mueller}@tu-dortmund.de

***Abstract*—This contribution provides a novel graph representation-based approach for the re-identification of chipwood surface structures and, in the herein observed use case, the re-identification of Euro-pallets. For this purpose, we suggest, in contrast to common re-identification approaches, replacing the usual image representation with a highly compressed graph representation. This allows for the creation of an efficient algorithm while also providing robustness to environmental changes such as rotation and shearing. The resulting method, called *IRAG* (Image Representation through Anomaly Graphs), is a siamese graph neural network, that is applied on a previously published dataset consisting of images from 502 EPAL pallet blocks. The results of this approach lead to a rank-1 accuracy of 27% when re-identifying pallet blocks. Even though *IRAG* does not yet reach accuracy values that are comparable to state-of-the-art literature, it is however more efficient, concerning the handling and representation of data. In addition, the experiments in this work demonstrate that the re-identification accuracy of the model is not affected by rotation or shearing, demonstrating the model's invariance to these environmental changes.**

***Index Terms*—graph representations, resource efficient neural networks, re-identification, warehousing logistics**

I. Introduction and Motivation

Euro-pallets are a widely used type of load carrier, with hundreds of millions of them being in constant circulation [1]. Since they are primarily moved and shipped while carrying goods, their location and identity represent domain knowledge that can be beneficial to operators that deal with the pallet in question. However, Euro-pallets, like other standardized load carriers, do not hold any inherent identification methods and are instead only identifiable by batches, defined by certain criteria, e.g., their place of assembly [2]. In this context, Euro-pallets can thus not be identified as individual pallets but rather classified by a given set of criteria. Therefore, the identification of load units commonly takes place through documentation (e.g., waybills) or artificial markers (e.g., QR-codes) that are attached and linked to the good itself. The motivation for this approach lies in the relative (in comparison to the goods) inexpensiveness of load carriers and the frequent change of the goods they carry. If however, it were technically and economically feasible to identify load carriers based on their inherent visual characteristics, while linking them to the good they are currently carrying, further knowledge could be gathered about the load unit and it would permit related processes to first be analyzed and subsequently to be improved upon. A first attempt to solve this problem was presented in [3], in which the authors demonstrated the feasibility of identifying the chipwood pallet blocks used in Euro-pallets based on their unique wooden patterns. In this contribution, the pallet block images were directly converted to feature vectors. While a vectorized representation already greatly reduces the dimensionality of data (compared to an image), the specific use case of load carrier re-identification necessitates a resource and time-efficient data handling process.

Therefore, in this work, we provide a novel, alternative approach, which aims at the reduction of data dimensionality, while preserving the information necessary for re-identification, and increasing robustness. The herein presented approach is based on graph representations of the parts of the images that are considered to be anomalous. These anomalies are represented as the graph's nodes, which are connected by edges that represent the distances between the nodes. The characteristic features of the respective pallet block surface, such as prominent chips (i.e., visually distinguishable by size, color, etc.), are exploited and, hinging on an efficient training procedure, detected as anomalies in the chipwood pattern. By representing this set of information as a graph, the dimensionality of the data is further reduced, while maintaining the key features of the images. At the same time, the representation as a graph ought to increase the robustness of the resulting re-identification method, making the representation invariant to different circumstantial characteristics and environmental recording conditions, such as rotation, lighting, etc.

The remaining publication is structured as follows: The related work concerning re-identification approaches and graph representation methods will be presented in the proceeding Section II. Subsequently, the methods used to put the approach described above into practice are presented in Section III. Section IV discusses the results gathered from the experiments conducted for this contribution, before Section V finally provides a recapitulation and an outlook for this work.

II. Related Work

This section will be divided into a subsection concerning the state-of-the-art approaches and general concept of re-identification, and a subsection focusing on the methods of graph representation.

*Corresponding authors. These authors contributed equally to this work.

A. Re-Identification

Re-identification can be defined as the attempt to retrieve a previously recorded subject of interest over a network of cameras [4]. More specifically, re-identification systems are supposed to match input (query) data with a database (gallery), containing identities captured from different cameras at different points in time [5]. Commonly, a list of identities, ranked by their similarity to the query data, would be generated from this matching task. Re-identification therefore distinguishes itself from classification by going beyond class adherence and instead identifies individual subjects in a given dataset.

Re-identification is frequently used for pedestrian surveillance [4], [6]–[9], while other applications include vehicle surveillance [10]–[12], material identification [3], [13]–[16] as well as animal identification [17]–[19]. Since the methods of re-identification are most commonly applied to pedestrians, different kinds of visual identification criteria have been the subject of research and can be distinguished between. As such, full body re-identification based on the overall appearance of a person is a prominent one, while facial re-identification (or recognition) [20] and gait-based pedestrian re-identification [6] are of interest too. For non-human, industrial entities however, only visual characteristics and domain knowledge can reasonably be taken into account. Even then, distinct ways of approaching the task of re-identification do exist.

Different taxonomies of re-identification have been defined [7], [21] and distinguish the task by different metrics, such as the learning approach, the assumption of an open set versus a closed set of subjects, etc. A common learning approach is deep metric learning, in which the similarity between subjects is mapped as a feature embedding (often in a vectorized form) so that a query identity ought to have a higher similarity to another instance of itself in the gallery than to other identities [7]. In this context, similarities are articulated as distances between identities in the embedding space and the way in which these distances can be calculated is through loss functions. Formulating re-identification as a multi-class classification task, in which each class holds one entity, [22] calculated class adherence using cross-entropy loss. An alternative, especially for siamese networks, is contrastive loss [23], which, in current research, is partially replaced by triplet loss, adding the advantage of minimizing intra-class distances while maximizing inter-class distances [7], [24]. Other authors yet again argue for the use of classification loss [25], showing just how many different approaches to re-identification have been explored in recent years.

Irrespective of the loss function one wishes to choose, several evaluation metrics are commonly used to compare the resulting re-identification models. One such metric is ranked accuracy, also known as top-n accuracy, which describes whether the correct query identity is retrieved as part of the first n matches in the gallery data [4], [7]. The motivation behind the use of ranked accuracy is that it provides insight into the severity of a wrong retrieval, as there is a practical difference for many use cases, between the second or the ninetieth match being the correct identity, if the first one is not. From the ranked accuracy results, the CMC (cumulative matching characteristics) curve can be generated by summing up the accuracy of each query identity divided by the overall amount of query identities, giving a comprehensive understanding of the model's performance [4], [7]. In addition, mean Average Precision (mAP) is commonly used since it indicates the extent to which correctly retrieved images are at the top of the ranked results, measuring the average retrieval performance of the model [4], [7].

Analogous to and based on the state-of-the-art approaches described in the paragraphs above, [3] proposed a Euro-pallet re-identification pipeline, exploiting the unique chip-wood pattern of pallet blocks. For this contribution, pallet blocks are detected through YOLO-based object detection. The resulting images are passed through a ResNet50 implementation of PCB_P4 (Part-based Convolutional Baseline) [26], using cross-entropy loss, to be stored as a feature vector for the subsequent task of re-identification under a closed-set assumption. The results of the contribution suggest that the surface structure of chipwood can be reliably exploited for the purpose of re-identification, using the dataset *pallet-block-502*, that was created for this very purpose.

Currently, common methodological approaches to re-identification work with limited amounts of data and subjects. Due to these limitations, the characteristics of the subjects in question can be stored in a vectorized form. While the condensation through feature-vector representation already reduces the resulting data volume, especially when using low dimensional vectors, it still leads to large amounts of data and relies on a sophisticated model to generate them. In addition, this approach to data representation makes the output sensitive to environmental recording conditions. These conditions might be and are not limited to, e.g., rotations, shearing, and lighting. Finally, methodologically speaking, some approaches are highly specific to their intended use case, limiting their potential for generalizability. The input resolution of PCB_P4 for instance, with 384×128 px, provides a vertical aspect ratio, which is tailored to the model's intended use as a pedestrian re-identification network.

B. Graph Representation Learning

As graphs are designed to store relations between subjects, nodes and edges are often used to represent and analyze non-trivial subjects. Examples range from medicine over particle physics to linguistics [27]–[29]. Because of this diverse range of uses, the method of graph machine learning has gained popularity in recent years [30].

Even for images, a graph representation can outperform ordinary convolutions [31], as graphs allow for the modeling of structures instead of pixels and to store these structures in only a few features. This can for instance be realized through superpixels [32], which are regions consisting of multiple similar pixels that are combined into one node. In contrast to the superpixel approach, this work does not focus on representing individual pixels and instead focuses on

determining regions containing relevant information. In person re-identification, the use of graph neural networks is novel but gaining popularity. [33], [34] employ graphs to improve the triplet loss used in a siamese network. Displaying a higher degree of similarity to the herein presented approach, [35] use graphs to represent a person by their joints. However, while doing so, the authors of [35] focus on extracting information from related structures in time. Given these circumstances and the specific use case, their representation graph construction is heavily optimized for the task of person re-identification.

III. Methodology

This section provides insight into the methodology used to gather the results of this contribution. The section is divided into subsections detailing the tasks of data preparation, model training, and testing and evaluation of the resulting model.

A. Data Preparation

The dataset used for the experiments of this contribution, *pallet-block-502*, which consists of $5,020$ RGB images of pristine, unbranded EPAL pallet blocks of which $2,510$ are used for the herein presented experiments. Every pallet block has been captured using two different cameras (to test lighting invariance) and five different perspectives: central (C), left-hand and right-hand side rotation (RL, RR), and left-hand and right-hand side shift (L, R). Therefore, the dataset contains images of 502 different pallet blocks (see Fig. 1).

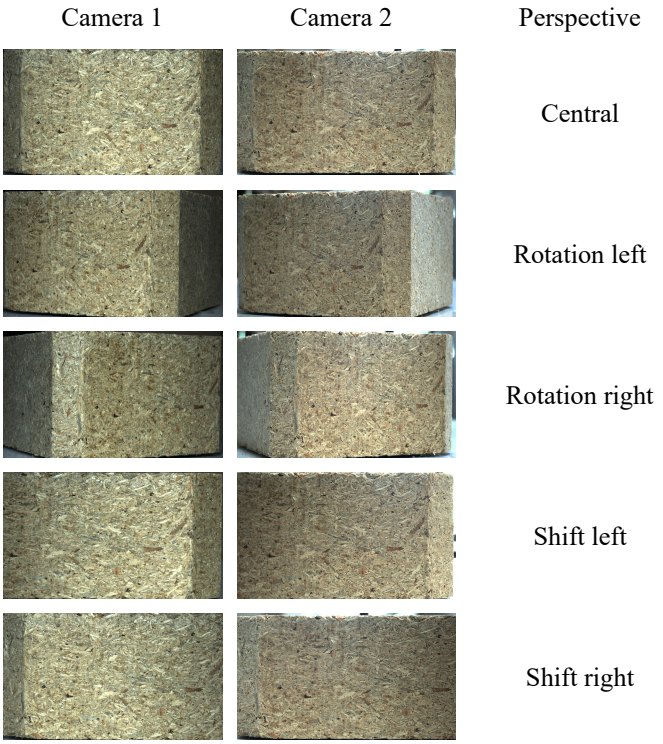

Fig. 1. The ten images given for one pallet block (ID 301).

Further details concerning the recording setup can be taken from [3] as well as from the resulting dataset, which is available online[1].

For this work, the dataset is used in the following way:

- 60% of the pallet blocks ($1,505$ images and 301 pallet blocks (and therefore IDs)) are used for training and validation;
- 20% of the pallet blocks (500 images and 100 pallet blocks) are used for testing;
  - One image per pallet block (100 images) is used as a gallery, the set of a priori known images under a closed-set assumption;
  - The remaining perspectives (400 images) of each pallet block in the gallery are used as a query set and are supposed to be matched with the gallery set;
- 20% of the pallet blocks (505 images and 101 pallet blocks) are held out for a future novelty detection task and are not used at this point in time.

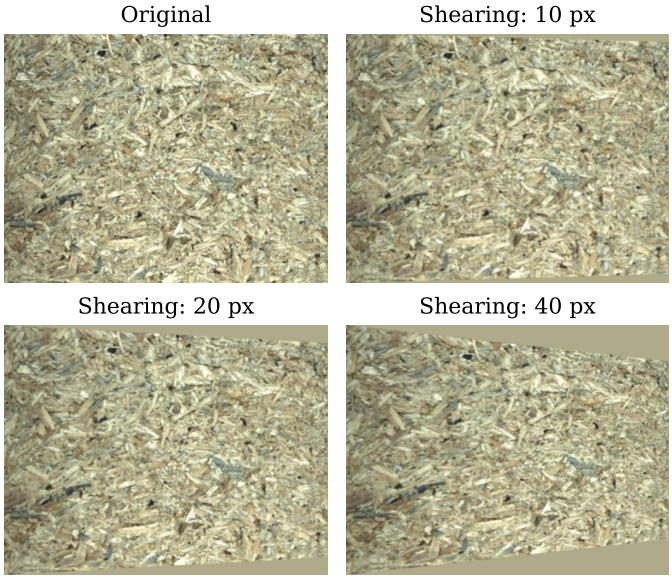

Fig. 2. The three chosen shearing degrees applied to one image (ID 306).

In addition, for the purpose of robustness experimentation, the dataset was augmented in the following way: using the Python library OpenCV, shearing was applied to the original dataset images. The authors assume that shearing is analogous to further rotation and distortion that might occur when observing pallet blocks from differing angles in realistic warehousing scenarios, e.g., when using on-board cameras on drones or autonomous guided vehicles. The representation of the data as a graph however, should make the re-identification results more invariant to shearing and rotation, compared to the representation in a vectorized form. Specifically, an effect that should mimic horizontal shearing is applied by mapping the rectangular image to a trapezoidal shape, in which the shorter side is smaller at both the bottom and the top of the image

[1] https://zenodo.org/record/6353714

(see Fig. 2). Shearing of varying degrees is used in this work, providing up to 2×40 px of length reduction on the short side of the trapezoid. The resulting padding that is added to the contours of the images is colored with the average RBG value of the dataset's images, as to not be interpreted as an anomaly by the herein presented model.

One further challenge that arises with the given dataset is that the chamfered edges of the pallet block (see Fig. 1) are distinct from the rest of the image, are not always in the camera frame and should thus generally not be considered to contain anomaly nodes. In order to remedy this wrongful consideration as anomalous, the images are cropped so that the chamfered edges of the pallet blocks are not visible. We use a simple heuristic, preferably removing too much of the image, rather than accidentally leaving a chamfered edge in the resulting dataset. This means that information is lost during pre-processing, which might further complicate the task at hand. At the same time, it might also make the resulting algorithm more robust, as pixels on the side of the center area are not always clearly visible (consider the right-hand side shift in Fig. 1). We subsequently rescale our images to a size of 500×300 px, roughly preserving the original average image aspect ratio of 1.7.

B. Methodological Approach

Our method can be subdivided into three key components, which are visualized in Fig. 3:

1) The information stored in the dataset images is transferred into a graph. This allows for the removal of a large portion of features from the dataset while retaining the information needed. Given this background, we assume that the parts of the pallet blocks that contain the largest amount of relevant information are the most visually striking pieces of chipwood. The anomalies in these image parts likely contain most of the information needed for re-identification.
2) We use highly reduced (in terms of dimensionality) graph representations to train a siamese neural network to further reduce the data dimensionality. During this process, we map images of the same pallet block to similar positions in the output space, while mapping different blocks to distinct embedding space positions.
3) We use an algorithm to match new images in their graph representation, i.e., the siamese network's output, to the a priori known image that is the closest in the embedding space (i.e., the most similar one). For this purpose, a k-nearest neighbor algorithm with $k = 1$ is used.

C. Graph Representation Procedure

Since we consider anomalies in the dataset images to be points of interest, it is sensible to define these anomalies as the nodes of the representation graph. This requires an efficiently working algorithm to find anomalies in a given image. We iterate over each 16×16 px area (with a 50% overlap between successive areas) of the image and calculate an anomaly score for it. As this requires calling the anomaly detector multiple times per image and we are presented with a narrowly defined use case, a simple anomaly detector is used. This anomaly detector considers a subregion (the 16×16 px areas described above) S to be more anomalous the higher $A(S)$ is (with the average $mean(S)$ and the standard deviation $std(S)$). $A(S)$ is defined as:

$$A(S) = \frac{std(S)}{mean(S)} \tag{1}$$

When using a simple algorithm like the one outlined in Equation 1, a challenge to be taken into consideration is the wrongful adjacency of nodes that display such extreme proximity, that they ought to be merged into a single node. In order to minimize this occurrence, we introduce a modification to the algorithm, in which all subregions are looped through, in descending order of $A(S)$. If the current S is close to another node ($mse(pos[curr], node) < \tau$), the average of the position of both nodes is calculated, merging both nodes into a single one. If S falls under this threshold, a new node is created. This process is repeated until a total of 50 nodes are generated, which we consider to provide a sufficiently complex graph, representing the given images accurately. This node-finding algorithm is described in Algorithm 1. In each node, information about a given anomaly and its surroundings is saved. A list of all parameters taken into consideration for each node can be found in Table I.

TABLE I
FEATURES REPRESENTING ONE NODE OF THE GRAPH.

Features		
Location [px]	Color [RGB]	Dimensionality
X-axis location	Average subregion color	Number of subregions S_i
Y-axis location	Color standard deviation	Subregion size

Subsequently, graph connections are generated using a top-k algorithm on the location of the anomaly, connecting each node with the five other nodes with the highest proximity. The aim of this method is to capture a minimalist version of the pallet block's surface structure as a graph representation.

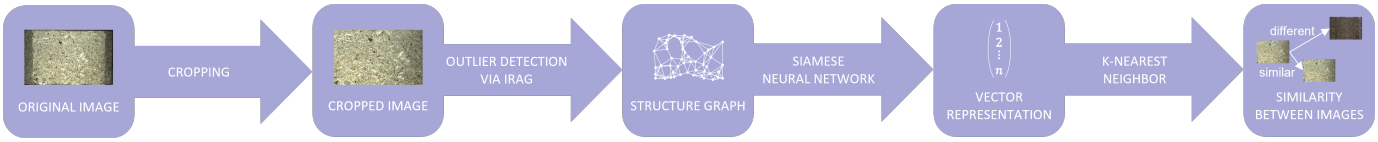

Fig. 3. Workflow developed in this contribution.

Algorithm 1 IRAG

Input: S_j, pos_j (subregions and their positions), $nodec$ (number of nodes to find), tau (minimum distance between two nodes)

Output: $nodes$ (locations of the nodes found)

$n \leftarrow 0$
$nodes_i \leftarrow (0,0)\ \forall i < nodec$
$regions_i \leftarrow 0\ \forall i < nodec$
$a_i \leftarrow A(S_i)\ \forall S_i$
$dex \leftarrow reverse(argsort(a_i))$
$k \leftarrow 0$
while $n < nodec$ **do**
  $curr \leftarrow dex[k]$
  $dist \leftarrow mse(pos_{curr} - nodes[j])\ \forall j < n$
  if $min(dist) < tau$ **then**
    $closest = argmin(dist)$
    $nodes_{closest} \leftarrow \frac{nodes_{closest} \cdot regions_{closest} + pos_{curr}]}{regions_{closest}+1}$
    $regions_{closest} \leftarrow regions_{closest} + 1$
  else
    $nodes_n \leftarrow pos_{curr}$
    $regions_n \leftarrow 1$
    $n \leftarrow n + 1$
  end if
  $k \leftarrow k + 1$
end while
return $nodes$

We show the features stored in each node in Table I and illustrate the node generation process in Fig. 4.

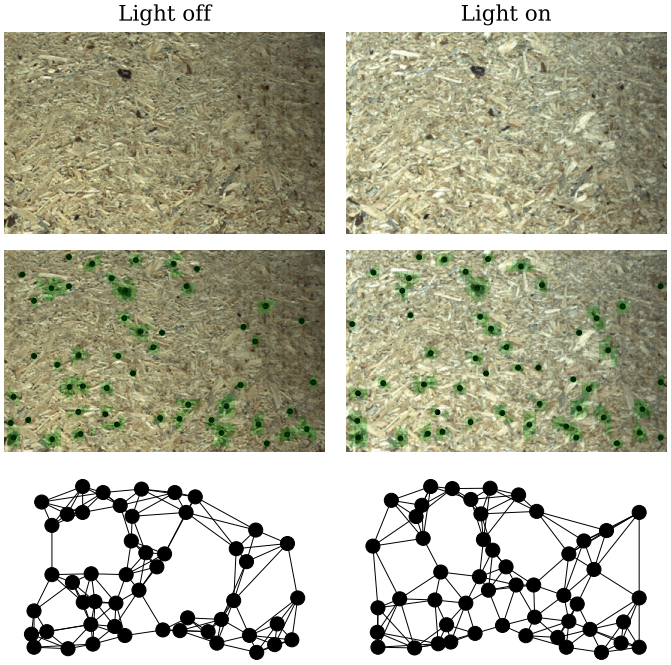

Fig. 4. Transforming an image of a pallet block (top row; ID 120) into a graph (bottom row) by means of anomaly detection (middle row)

D. Siamese Graph Neural Network Training

We train the herein proposed siamese graph neural network using triplet loss as a loss function. This means that each input consists of three samples: a comparison sample X_i, a sample of the same pallet block T_i, and one of a different pallet block F_i. Our model $r(X)$ is trained so that the loss function L_T is minimized:

$$L_T = \sum_i RL(D(r(X_i), r(T_i)) - D(r(X_i), r(F_i)) + \alpha) \quad (2)$$

Here, D represents the mean squared difference and the ReLU function is defined as $RL(x) = \frac{|x|+x}{2}$. x Intuitively, this approach aims at minimizing the embedding space distances between the same pallet block and maximizing the distance between different pallet blocks. We train the model described in Table II in batches of 100 images for up to 100 epochs with a patience of 5, using the Adam optimizer and a learning rate of 0.0001. We define the shape of each graph layer as $nodes \times features + nodes \times nodes$ (node features and adjacency matrix).

These graph convolutions are provided by the Python library grapa [36]: Each convolution combines the node values with the values of the nodes that they are connected to, using learnable weights. Subsequently, we combine the values of each node using a global pooling operation.

TABLE II
ARCHITECTURE OF THE SIAMESE GRAPH NEURAL NETWORK.

Layer type	Output shape	Activation function
Input	50 x 10 + 50 x 50	-
Graph Convolution	50 x 31 + 50 x 50	ReLU
Graph Convolution	50 x 31 + 50 x 50	ReLU
Global Mean Pooling	31	-
Fully Connected	250	ReLU
Fully Connected	250	ReLU
Fully Connected	250	ReLU
Fully Connected	50	Linear

E. Testing & Evaluation

As to test the performance of the proposed method, the model is applied to a hold-out dataset, split into gallery and query images, as previously described. The results are further subdivided by analyzing the model's performance regarding the use of different cameras as well as different perspectives. The metric of choice, used to evaluate the performance of the model is ranked accuracy (rank 1 to rank 10).

As to evaluate the degree to which shearing affects the performance of the model, the achieved rank-1 accuracy is compared over different degrees of shearing, as described in Section III-A (see Fig. 2). We assume that the resulting accuracy values follow binomial distributions and thus we approximate the uncertainty by the uncertainty of this distribution as $\sqrt{\frac{p \cdot (1-p)}{N}}$ (with p being accuracy and N the number of samples).

IV. Experimental Results

In this section, the experimental results of this contribution are presented. First, the results concerning the re-identification task are presented. Subsequently, the results concerning the robustness and circumstance invariance of our proposed method are shown.

A. Re-Identification Results

Even though our results are preliminary in nature, we achieve a performance that clearly distinguishes our model from a randomly guessing one.

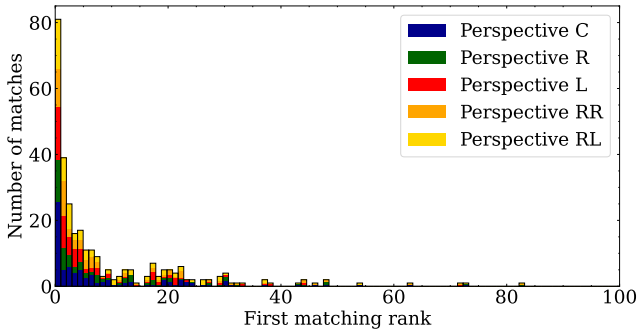

Fig. 5. Histogram of the first rank in the feature space at which a match is found, subdivided by the perspectives used per match.

Fig. 5, for instance, shows the number of gallery images that are closer to a query image than the comparison sample of the correct image. Using a perfectly functioning model, the diagram would be reduced to a column at 0 on the x-axis. Using the model presented in this contribution, the most dominant peak of the diagram still resides at 0. Still, some query samples are hardly identified correctly and one sample even seems more likely to be one of 83% of all other samples, than the one of the matching pallet block.

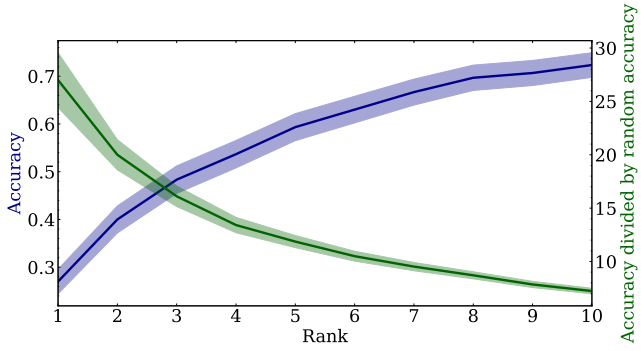

Fig. 6. Ranked accuracy of the re-identification task plotted as a CMC curve.

This can further be quantified using a CMC (cumulative matching characteristics) curve. We plot the rank-k accuracy against k in Fig. 6. The resulting rank-1 accuracy of 27% is not yet competitive against comparable methods, providing, e.g., $> 72\%$ rank-1 accuracy in [3]. Nevertheless, with the results of these experiments being preliminary, we provide novelty through the methodology itself and demonstrate the feasibility of graph representation-based re-identification. It should also be taken into account that on no occasion, the accuracy provided by our model results in random guessing, as the green curve in Fig. 6, even with uncertainty, is never equal to the null hypothesis of 1.

B. Efficiency

Compared to the $45,000$ features of the original images, we use 900 times fewer features in our graph representation (500 features). This translates to a much more memory and processing time-efficient computation. Due to this, we achieve a training duration of < 10 min compared to multiple hours of training when using images. It is also much cheaper to store these graph representations. After using a common compression, our training set has a size of 581 MB in the original image representation, while the graph representation only requires 3.7 MB. This is a factor of 157, as the images are more susceptible to compression.

C. Robustness Tests

The use of graphs instead of images ought to make our approach more robust. We have conducted three experiments to test this hypothesis: First, we notice that when using rotated versions of an image, the resulting graph representation remains the same. When applying a shearing effect to the images, similar albeit not equal results are to be expected. Shearing is not an effect that graphs are entirely invariant to, as these compress node positions on one side, which changes the distance matrix needed for the top-k algorithm generating our graphs. The results of the experiments conducted concerning shearing effects can be seen in Fig. 7. The results demonstrate the application of a model that was trained on images without shearing effects on sheared test images, analogous to the ones shown in Fig. 2. By plotting the relative rank-k accuracy against the shearing degree, we see an almost flat curve, implying the robustness of the developed model against shearing.

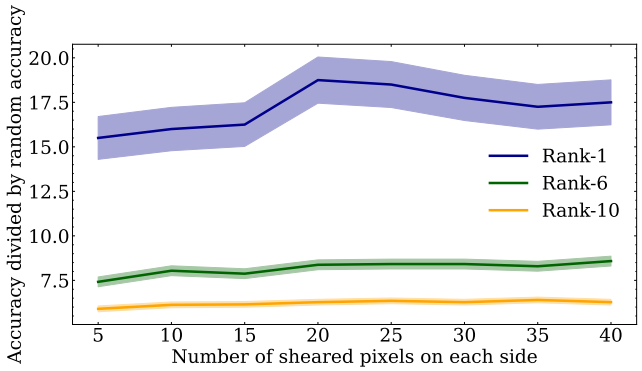

Fig. 7. Accuracy as a function of the shearing effect.

As a second test, we compare the re-identification results on the different kinds of images in the dataset (i.e., the different camera angles and perspectives used). A non-robust model might be better at re-identifying images of a certain type of

perspective over another. We show the results of testing this hypothesis in Fig. 8 by plotting the rank-k accuracy for each re-identification task that involves a certain perspective type of image (either as a gallery or a query image). As the changes in the resulting accuracy are limited, taking uncertainties into account, we can again demonstrate the model's robustness against changing perspectives.

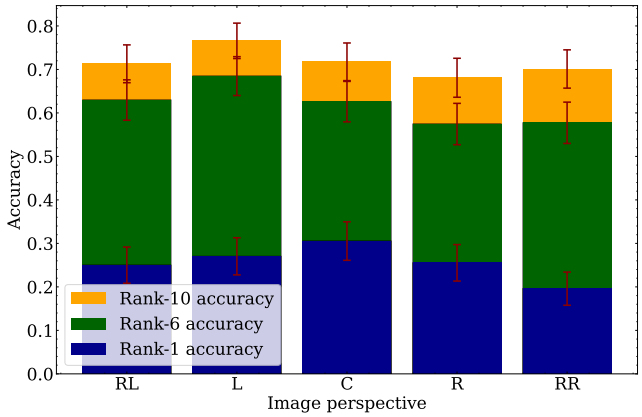

Fig. 8. Rank-1, rank-6 and rank-10 accuracy with their corresponding uncertainties, ordered by perspective.

Finally, we exploit the fact that the dataset contains images of the same pallet blocks and perspectives, captured by two different cameras. As the graph representation should not be affected by changes in blur or brightness (up to node features), we assume that after training on images taken by one camera, the resulting model should provide a similar accuracy when tested on images recorded by the second camera. This would be quite useful, as in practice, it is not feasible to retrain the model every time a new camera is used.

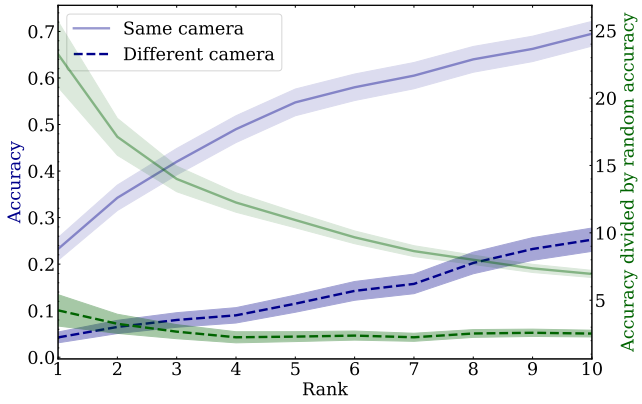

Fig. 9. Comparing CMC curves for different cameras.

To our dismay, as Fig. 9 shows, invariance to camera types cannot yet be demonstrated. We assume this problem to stem from the graph generation process and consider it future work.

V. Conclusion & Outlook

In this contribution, the authors provided a novel graph representation-based approach for the re-identification of chipwood pallet blocks. Instead of using convolutional image neural networks, the authors suggest using anomalies in the structure of the pallet blocks to build a graph out of an image and use graph neural networks to re-identify query images. The herein presented method, called *IRAG*, was applied on a previously published dataset, *pallet-block-502*, consisting of $5,020$ images of chipwood pallet blocks, belonging to Europallets. The images of this dataset are taken under varying angles and with different cameras, providing a certain degree of heterogeneity to the data. The application of the herein presented method on the dataset resulted in a rank-1 accuracy of 27%, proving the feasibility of the concept, while not being ready for industrial usage yet. Beyond proving the feasibility of our approach, it was also shown, that the representation of data as a graph leads to an invariance regarding rotation and shearing. An invariance concerning lighting was, contrary to the authors' assumptions, not demonstrated through the experiments of this work. In addition, the certainty with which the re-identification task was performed still leaves room for improvement and optimization.

Our key goal for future research is to improve the quality of the results of our experiments. This primarily means reaching a competitive accuracy. We have currently identified four ways in which this could be achieved:

1) A more precise image cropping algorithm as a preprocessing step before the node generation will be needed, since the current heuristic tends to remove more information than necessary.
2) Currently, $A(S)$ is biased against large white spaces and thus should further be tuned.
3) Grapa is a fairly rudimentary GNN library and replacing it with a more commonly used one might be helpful. This might also allow for more precise structure encoding, e.g., by adding information to the graph's edges.
4) The performance of the siamese graph neural network can benefit from further hyperparameter tuning.

In addition, the developed method could be applied to other datasets, containing similar images, as to test the extent to which the results and the approach are generalizable. By changing the definition of anomalous regions, for instance, our approach could be deployed on other datasets. Finally, provided with a more general definition of what contributes to anomalous regions, the algorithm might also be able to provide users with more trustworthy re-identification results.

Acknowledgment

This work is part of the project "Silicon Economy Logistics Ecosystem" which is funded by the German Federal Ministry of Transport and Digital Infrastructure.

The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

This work was supported by the Research Center Trustworthy Data Science and Security, an institution of the University Alliance Ruhr.

This research has been funded by the Federal Ministry of Education and Research of Germany as part of the competence center for machine learning ML2R ($01IS18038$)

REFERENCES

[1] I. Deviatkin and M. Horttanainen, "Carbon Footprint of an EUR-sized Wooden and a Plastic Pallet," in *E3S Web of Conferences*, vol. 158. EDP Sciences, 2020.

[2] D. S. C. Packaging, "Pallet Production Specification - Part 1: Construction Specification for 800 mm × 1200 mm Flat Wooden Pallets; German version EN 13698-1:2003," Beuth Verlag, Tech. Rep., 2004.

[3] J. Rutinowski, T. Chilla, C. Pionzewski, C. Reining, and M. ten Hompel, "Towards Re-Identification for Warehousing Entities – A Work-in-Progress Study," in *Proceedings of the IEEE Conference on Emerging Technologies In Factory Automation (ETFA)*, 2021, pp. 501–504.

[4] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep Learning for Person Re-Identification: A Survey and Outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2872–2893, 2021.

[5] E. Yaghoubi, A. Kumar, and H. Proença, "SSS-PR: A Short Survey of Surveys in Person Re-Identification," *Pattern Recognition Letters*, vol. 143, pp. 50–57, 2021.

[6] A. Nambiar, A. Bernardino, and J. C. Nascimento, "Gait-Based Person Re-Identification: A Survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–34, 2019.

[7] Z. Ming, M. Zhu, X. Wang, J. Zhu, J. Cheng, C. Gao, Y. Yang, and X. Wei, "Deep Learning-Based Person Re-Identification Methods: A Survey and Outlook of Recent Works," *Image and Vision Computing*, vol. 119, p. 104394, 2022.

[8] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, and D.-S. Huang, "Deep Learning-Based Methods for Person Re-Identification: A Comprehensive Review," *Neurocomputing*, vol. 337, pp. 354–371, 2019.

[9] Q. Leng, M. Ye, and Q. Tian, "A Survey of Open-World Person Re-Identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1092–1108, 2019.

[10] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-Scale Vehicle Re-Identification in Urban Surveillance Videos," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2016, pp. 1–6.

[11] L. Wei, X. Liu, J. Li, and S. Zhang, "VP-ReID: Vehicle and Person Re-Identification System," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, 2018, pp. 501–504.

[12] S. D. Khan and H. Ullah, "A Survey of Advances in Vision-Based Vehicle Re-Identification," *Computer Vision and Image Understanding*, vol. 182, pp. 50–63, 2019.

[13] A. Klar, "Identifizierung von Ladehilfsmitteln auf der Basis Stochastischer Oberflächeninformationen," Ph.D. dissertation, TU Dortmund University, Germany, 1995.

[14] T. Takahashi, Y. Kudo, and R. Ishiyama, "Mass-Produced Parts Traceability System Based on Automated Scanning of "Fingerprint of Things"," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, 2017, pp. 202–206.

[15] T. Takahashi and R. Ishiyama, "FIBAR: Fingerprint Imaging by Binary Angular Reflection for Individual Identification of Metal Parts," in *Proceedings of the Fifth International Conference on Emerging Security Technologies (EST)*, 2014, pp. 46–51.

[16] J. C. Hermanson and A. C. Wiedenhoeft, "A Brief Review of Machine Vision in the Context of Automated Wood Identification Systems," *IAWA Journal*, vol. 32, no. 2, pp. 233–250, 2011.

[17] N. Lokare, Q. Ge, W. Snyder, Z. Jewell, S. Allibhai, and E. Lobaton, "Manifold Learning Approach to Curve Identification with Applications to Footprint Segmentation," in *2014 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*, 2014, pp. 1–8.

[18] Z. Jewell, S. Alibhai, F. Weise, S. Munro, M. Vuuren, and R. Vuuren, "Spotting Cheetahs: Identifying Individuals by their Footprints," *Journal of Visualized Experiments*, vol. 2016, 2016.

[19] B. V. Li, S. Alibhai, Z. Jewell, D. Li, and H. Zhang, "Using Footprints to Identify and Sex Giant Pandas," *Biological Conservation*, vol. 218, pp. 83–90, 2018.

[20] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face Recognition Systems: A Survey," *Sensors*, vol. 20, no. 2, 2020.

[21] N. K. S. Behera, P. K. Sa, S. Bakshi, and R. P. Padhy, "Person Re-Identification: A Taxonomic Survey and the Path Ahead," *Image and Vision Computing*, vol. 122, p. 104432, 2022.

[22] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person Re-Identification: Past, Present and Future," *arXiv preprint: 1610.02984*, 2016.

[23] R. R. Varior, M. Haloi, and G. Wang, "Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification," in *European Conference on Computer Vision*. Springer, 2016, pp. 791–808.

[24] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person Re-Identification in the Wild," *arXiv preprint: 1604.02531*, 2016.

[25] Y. Zhai, X. Guo, Y. Lu, and H. Li, "In Defense of the Classification Loss for Person Re-Identification," *arXiv preprint: 1809.05864*, 2018.

[26] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline)," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 480–496.

[27] F. Agosta, S. Sala, P. Valsasina, A. Meani, E. Canu, G. Magnani, S. F. Cappa, E. Scola, P. Quatto, M. A. Horsfield, A. Falini, G. Comi, and M. Filippi, "Brain Network Connectivity Assessed Using Graph Theory in Frontotemporal Dementia," *Neurology*, vol. 81, no. 2, pp. 134–143, 2013.

[28] R. P. Feynman, "Space-Time Approach to Quantum Electrodynamics," *Physical Review Journals*, vol. 76, pp. 769–789, 1949.

[29] A. Piperski, "An Application of Graph Theory to Linguistic Complexity," *Yearbook of the Poznan Linguistic Meeting*, vol. 1, 2014.

[30] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph Neural Networks: A Review of Methods and Applications," *arXiv preprint: 1812.08434*, 2018.

[31] P. H. C. Avelar, A. R. Tavares, T. L. T. da Silveira, C. R. Jung, and L. C. Lamb, "Superpixel Image Classification with Graph Attention Networks," *arXiv preprint: 2002.05544*, 2020.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[33] D. Chen, D. Xu, H. Li, N. Sebe, and X. Wang, "Group Consistent Similarity Learning via Deep CRF for Person Re-identification," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8649–8658.

[34] Y. Shen, H. Li, S. Yi, D. Chen, and X. Wang, "Person Re-Identification with Deep Similarity-Guided Graph Neural Network," *arXiv preprint: 1807.09975*, 2018.

[35] K. Khaldi, P. Mantini, and S. K. Shah, "Unsupervised Person Re-Identification Based on Skeleton Joints Using Graph Convolutional Networks," in *Image Analysis and Processing – ICIAP 2022*. Springer International Publishing, 2022, pp. 135–146.

[36] S. Klüttermann, "grapa," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.7185762