ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

# Unpaired Image-to-Image Translation to Improve Log End Identification

Dag Björnberg[1,2], Morgan Ericsson[1], Welf Löwe[1,2] and Jonas Nordqvist[1] *

1- Linnæus University, Universitetsplatsen 1 - Växjö, Sweden

2- Softwerk AB, Reveljgränd 5 - Växjö, Sweden

**Abstract**. Visual re-identification tasks are often subject to large domain variations due to camera types, brightness conditions, or environmental differences. For identification models to generalize in such varying domains, a large amount of training data is necessary for capturing these variations. We explore the potential of using unpaired image-to-image translation to enhance the generalization capacity of a log end identification model in the absence or combined with a smaller amount of labeled training data.

## 1 Introduction

Deep learning is used in many state-of-the-art machine vision systems. However, their domain adaptation capabilities are often limited. For re-identification systems, in particular, the source data can be varying due to camera types or brightness and illumination conditions, causing a significant drop in performance when such a system is exposed to new variants of the data [1, 2]. Moreover, supervised identification models rely on paired data, *i.e.*, positive matches between images that are potentially taken under different conditions. The pairs must be collected or verified manually, which is a time-consuming process.

To address data variation and the expensive labeling of positive matches, we investigate whether unpaired image-to-image translation can be utilized to improve the generalization capabilities of an identification model. We apply and test this idea in the concrete context of tree log matching based on log end surface images. More specifically, we examine whether images captured by a mobile device in open-air surroundings can be recognized in a sawmill environment. By using unpaired image-to-image translation, we demonstrate significant improvements regarding the generalization capabilities of our identification model in this new environmental setting.

To perform unpaired image-to-image translation, we use CycleGAN [3], which has been successfully applied to many unpaired image-to-image translation tasks [4, 5]. In short, given two datasets $\{x_i\}_{i=0}^N$ and $\{y_i\}_{i=0}^M$ from domains $A$ and $B$, CycleGAN learns two mappings $G\colon A \to B$, and $F\colon B \to A$, such that $G(A)$ is indistinguishable from $B$ and $F(B)$ is indistinguishable from $A$. Both generators are trained in an adversarial fashion, with two discriminators $D_A$ and $D_B$,

*DBs work was funded by the Industry Graduate School on “Data Intensive Applications (DIA)” at Linnæus University which is partially funded by the Knowledge Foundation (project id 20190336).

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

learning to separate generated samples from $G(A)$ and $F(B)$ from real samples from $B$ and $A$, respectively.

## 2 Background

Tracking timber throughout the supply chain is an important mechanism to prevent illegal logging [6, 7]. Consequently, several traceability systems have been proposed. Many of these systems are based on non-automated labeling approaches such as attaching barcodes, QR codes, or RFID transponders to logs [6]. They are thus limited by their manual nature, making them ineffective, costly, and prone to human error. Approaches based on biometric characteristics of log end faces have been proposed to combat such limitations.

We have developed an identification model based on convolutional neural networks, similar to the method described in [8], to recognize log ends between two stations in a sawmill environment: sorting station and sawmill station. The identification model maps every log end image from the sawmill station to the closest match from the sorting station from the preceding 60 days. In this production environment, we have reached a correct identification rate of around 95%.

For complete traceability, the timber must be traced back to its origin, the harvesting site. This involves mapping every log from the sorting station to the correct match from the harvesting site. To assess the robustness of our identification model in such a future scenario, we conducted tests using mobile images of log ends captured in open-air surroundings. When the logs passed the sorting station, we checked how many that were correctly identified and noticed a significant drop in accuracy, down to around 70%, reflecting deficiency in identification systems when exposed to new domains. With the large number of sources in timber-based industries, all with potential variations in lightning conditions, cameras, and wood species, we are encouraged to investigate how we can bridge the domain gap from the existing identification model to a new environmental setting. As we strive for complete traceability in the timber supply chain, the ability to generalize our identification model to new circumstances is crucial, and a solution to the discovered problem is necessary.

## 3 Related Work

We are not the first to apply CycleGAN when dealing with images from different camera domains or environmental settings for identification tasks. CycleGAN and variants thereof have been employed extensively in various person re-identification tasks. For instance, it has been used to tackle the issue of cross-camera domain variation [1, 2], and to deal with the class-imbalance of RGB and IR images [9].

Although our work leverages CycleGAN, alternative approaches within the realm of unpaired image-to-image translation can also be applied within our framework, such as multimodal unsupervised image-to-image translation [10], contrastive learning [11], or StarGAN [12]. These methods offer complemen-

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

tary approaches for potentially improving re-identification of log ends. However, testing these alternative strategies for our purposes is left as future work.

## 4 Methodology

Approximately 6 500 matches have been collected to train the identification model between the sorting station and the sawmill station. We also have access to about 4 500 images taken by a mobile device in open-air surroundings. In addition, 492 matches between mobile images and the sorting station were collected to evaluate the performance of our suggested approaches.

### 4.1 Implementation and Initial Experiments

For training of the CycleGAN, we collect a combined sample of 4 510 images from the sorting and sawmill stations to match the number of mobile images. The images were resized to $512 \times 512$. The batch size was set to 1, as in [1], and we trained for 3 epochs for a total of 13 500 training steps. The generator architecture is a U-Net with skip connections, as described in [13]. The discriminator architecture is a convolutional neural network with downsampling layers.

When training the identification model, we perform live augmentations, meaning that new augmentations are performed between epochs. Such augmentations include rotations, perspective transformations, random cutouts, and contrast variations. These augmentations were used during training of our baseline model for the sorting-sawmill matches. To generalize our model to the new setting with mobile-sorting matches, we enhance standard augmentation techniques by additionally converting 25% of the sorting-sawmill images to mobile style during training. This decision is based on [2, 9], where a 3:1 ratio between the real and synthetic (style-transferred) images was found to be optimal.

The bidirectionality of CycleGAN encourages us to investigate yet another approach. Instead of converting sorting-sawmill images to mobile style during training, we examine if we can perform the reverse transformation and convert the mobile images to sorting-sawmill style before matching them with the sorting images. With this approach, no retraining of the identification is needed; instead, the transformation is used as a preprocessing. See Figure 1 for examples of style-transferred images and Figure 2 for a schematic overview of the two approaches.

### 4.2 Experimental Setup and Evaluation

Assessing the models' performance in the production environment consists of a deployment phase in combination with a manual verification of the correct match percentage. This is a time-consuming process, so we constructed a "virtual production line". This involved collecting an expanded set of 20 000 images from the sorting station in addition to the 492 matches. For each mobile image, we mapped it to its closest match out of the $20\,000 + 492$ images from the sorting station. We noticed a baseline accuracy of around 70%, consistent with initial observations from production. Consequently, we employed this methodology to

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

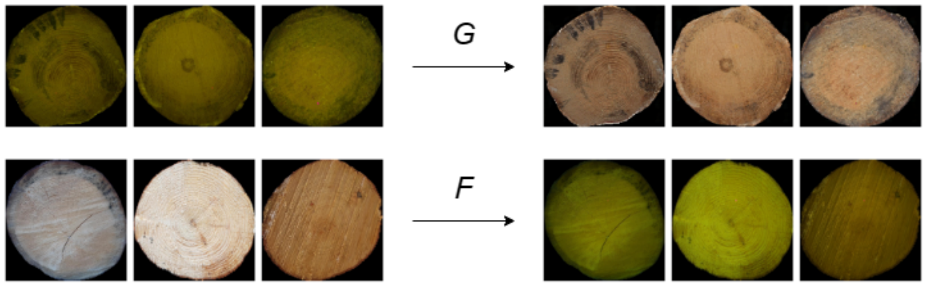


Fig. 1: Examples of style-transferred images using CycleGAN. The generator $G$ produces mobile stylized images from sorting-sawmill images, while the generator $F$ converts mobile images to sorting-sawmill style.

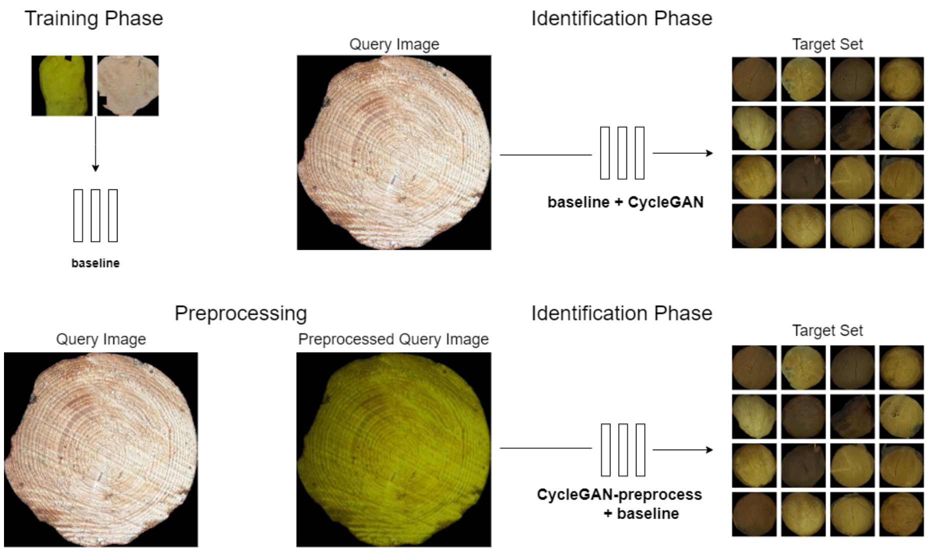


Fig. 2: Schematic overview of two approaches. The baseline model is retrained with style-transferred images in the first approach (top). The second approach (bottom) uses the reverse transformation as a preprocessing before it is mapped to its closest match from the sorting images.

evaluate our approach, serving as a proxy for accuracy in production. In addition to top1-accuracy, we also measured the top5-accuracy.

After the initial experiments, we performed additional experiments in which real mobile-sorting matches were included in the training phase. This was achieved by including 412 of the 492 mobile-sorting matches in the training phase of the identification model, leaving the remaining 80 matches for evaluation. Additionally, another 85 mobile-sorting matches were collected and added for evaluation purposes. To summarize, 412 mobile-sorting matches were used

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

during the training phase, and 165 matches were used for evaluation purposes. The identification model was trained with and without style-transferred images to assess whether or not CycleGAN can be used to boost the performance of our identification model when real mobile-sorting matches are available.

## 5 Results

From the initial experiments, we were able to improve the baseline accuracy. This was achieved in two ways: using CycleGAN style-transferred images during training and the reverse transformation as a preprocessing; see Table 1.

| **Model** | **top1-acc. (%)** | **top5-acc. (%)** |
|---|---:|---:|
| baseline | 70.1 | 82.3 |
| baseline + CycleGAN | 84.5 | 92.3 |
| CycleGAN-preprocess + baseline | 85.7 | 93.3 |

Table 1: Comparison of different methods regarding accurate predictions of 492 mobile-sorting matches.

Continuing our investigation, we decided to include real mobile-sorting matches in the training phase of the identification model. We performed two training iterations, with and without CycleGAN style-transferred images. The results can be seen in Table 2.

| **Model** | **top1-acc. (%)** | **top5-acc. (%)** |
|---|---:|---:|
| baseline | 72.1 | 82.4 |
| baseline + real matches | 92.1 | 97.0 |
| baseline + real matches + CycleGAN | 98.2 | 100.0 |

Table 2: Comparison of different methods regarding accurate predictions of 165 mobile-sorting matches.

## 6 Conclusion

The research conducted in this paper originates from the inherent uncertainty of machine learning models once exposed to data from new domains. Our investigation aimed to examine this on an identification model of log ends, where we noticed a significant drop in performance when images from a new source were added. To bridge this gap, we propose using CycleGAN for image translation.

We assessed whether applying CycleGAN solely could improve our identification model in two ways: by adding style-transferred mobile images during training and by using the reverse transformation as a preprocessing. In the first approach, style-transferred mobile images are added during the training phase, which can lead to model generalization. The second approach converts mobile images to sorting-sawmill style as preprocessing, which might be preferred in situations where retraining is infeasible or expensive. For both approaches, we noted a similar increase in accuracy, where top1-accuracy improved from 70.1% to around 85%.

After the initial experiments, we investigated whether style-transferred images can improve performance when real training data is available. This was achieved by including real mobile-sorting matches during training. To assess whether performance can be improved, we trained the identification model twice. First by using real training data only, we achieved a top1-accuracy of 92.1%. Second, by using a combination of real training data and style-transferred images, we further improved accuracy, with a top1-accuracy of 98.2%.

## References

[1] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camera style adaptation for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5157–5166, 2018.

[2] Zhenzhen Yang, Jing Shao, and Yongpeng Yang. An improved cyclegan for data augmentation in person re-identification. *Big Data Research*, 34:100409, 2023.

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[4] Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):16884, 2019.

[5] Wansik Choi, Jun Heo, and Changsun Ahn. Development of road surface detection algorithm using cyclegan-augmented dataset. *Sensors*, 21(22):7769, 2021.

[6] Ioakeim k Tzoulis, Zacharoula S Andreopoulou, and Elias Voulgaridis. Wood tracking information systems to confront illegal logging. *Journal of Agricultural Informatics*, 5(1), 2014.

[7] Rudolf Schraml, Johann Charwat-Pessler, Alexander Petutschnigg, and Andreas Uhl. Towards the applicability of biometric wood log traceability using digital log end images. *Computers and Electronics in Agriculture*, 119:112–122, 2015.

[8] Georg Wimmer, Rudolf Schraml, Heinz Hofbauer, Alexander Petutschnigg, and Andreas Uhl. Two-stage CNN-based wood log recognition. In *International Conference on Computational Science and Its Applications*, pages 115–125. Springer, 2021.

[9] Daoxun Xia, Haojie Liu, Lili Xu, and Linna Wang. Visible-infrared person re-identification with data augmentation via cycle-consistent adversarial network. *Neurocomputing*, 443:35–46, 2021.

[10] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, pages 172–189, 2018.

[11] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, pages 319–345. Springer, 2020.

[12] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.