Fully Automated, Realistic License Plate Substitution in Real-Life Images

Ufuk Kacmaz[1]   Jan Melchior[1]   Daniela Horn[2]   Andreas Witte[1]   Sebastian Schoenen[1]   Sebastian Houben[2]

Abstract— **Data augmentation techniques have been focused in recent research as they hold the promise to reduce the need for extensive data acquisition and to enable systematic sampling, e.g., in order to examine underrepresented cases. The question of how and to what extent control over the result is possible and necessary is still open.**

We propose a novel system for license plate substitution in the wild, which replaces a given license plate within an image crop by another one with a customized pattern. The system is based on a CycleGAN architecture, which respects the plate's pose and dominant image features, such as lighting and image sharpness. Most importantly the system is trained on a set of license plate crops without requiring any label information.

We demonstrate the validity of our approach in two sets of experiments: First, a license plate recognition system is trained and evaluated with varying amounts and ratios of synthetic over real-life data, and second, the realism of image features is verified by means of a human acceptance study as well as the Fréchet Inception Distance.

I. INTRODUCTION

Deep neural networks form an integral part of autonomous driving as their performance in solving sensor processing problems is unmatched. However, as the training of these models is almost exclusively data-driven, a demand for large and meticulously annotated datasets arises. Composing these datasets remains a laborious and costly task and means for its automatization have been extensively studied.

A promising approach is perceived in data augmentation techniques that aim to generate realistic data and annotations within given specifications. Apart from reducing the general amount of labor, they yield reliable annotations in complex scenarios while errors in manual labeling become more frequent. Furthermore, they can avoid recollecting and reprocessing data in case of a distributional shift. Examples for this phenomenon include the extension of the operational design domain, e.g., if the system is to operate in another country, and temporal domain shifts, which occur, for example, if the data correlate to trends and fashions. Depending on the generation system, partial control over the results can be granted such that certain sources of variance can be regulated or sampling can be performed evenly along given semantic dimensions. In consequence, it is possible to extend the same image material systematically, e.g., for various types of weather conditions or any time of day or night. Deploying a, in this sense, complete dataset then allows for systematic

[1] ControlExpert GmbH, 40764 Langenfeld, Germany `j.melchior@controlexpert.com`
[2] Institute for Neural Computation, Ruhr University Bochum, 44801 Bochum, Germany `daniela.horn@ini.rub.de`

Fig. 1. Generated license plate samples. The proposed system is able to replace existing license plates with synthetic ones while, at the same time, preserving perspective, lighting condition, and sharpness of the original real-life image.

evaluation and testing or for the removal of an assumed bias from the data.

In this paper, we investigate the problem of data augmentation for image-based license plate recognition, i.e., we evaluate automatically generated images of German vehicle registration plates and compare them to real-life samples. This is executed by an automated system performing a character-by-character extraction of the license plate information on a given image patch that has been obtained in a previously conducted detection process. The image generation pipeline is extendable and yields visually convincing, life-like results without the need for any label information. Fig. 1 illustrates exemplary images generated by our system.

Due to the simple geometry of license plates, it is straightforward to manage some properties of the real-life images that underlie elementary rules, such as perspective rendering or regulations according to which valid license plate numbers are chosen. In contrast, the properties related to appearance are learned from data by use of a style transfer model.

Our contributions in this paper can be summarized as follows:

- We present and describe in technical detail the data generation process and the ways in which it enables the user to control the outcome,
- we provide a comprehensive analysis with the help of a fully-fledged license plate recognition system using systematically generated data from our approach, and
- we evaluate the realism of the created results by means of a user acceptance study and the Fréchet Inception Distance.

II. RELATED WORK

Life-like automated image substitution is an active field of research as many data-driven models and applications

based thereon can benefit directly from better performance at a fraction of the annotation cost. It is to be juxtaposed with style transfer techniques, which generate images from a purely synthetic input, and object insertion, which embeds new objects into an image showing an otherwise empty scene. Depending on the complexity of the created scene, with style transfer, the gap between generated and real data can be anything from negligible [1] to severe [2], [3]. Object insertion aims to avoid this distributional shift by reusing real-life image data. Nevertheless, a new challenge arises from the need for a real-life embedding which aligns well with the surrounding and the global properties of the given scene [4].

Image substitution circumvents this demand to some extent by using objects that are already present in the scene and exchanging only key features. However, depending on the type of object or image feature one aims to modify, additional annotations might be necessary to maintain geometrical and visual consistency. Alternatively, one may deploy a potent detector or pose estimator, which in turn has to be trained from these annotations.

When addressing face swapping, most approaches rely on positions for facial landmarks to insert a face with a fitting pose and expression. Widely available annotated data [5] as well as public and commercial interest drive the research in this field. Ren et al. [6] propose a system for adversarial face swapping with two goals: minimum probability of re-identification and maximum performance for action recognition. An investigation of a close-to-market system was presented last year by Sümer et al. [7], in which the authors assess a real-time image and audio processing system for the anonymization of video calls in a challenging classroom scenario. Here, detected faces are, however, automatically blurred and not substituted. A recent approach by Hukkelås et al. [8], termed DeepPrivacy, aims at replacing faces without altering the position of facial landmarks in the process.

One of the main applications for these approaches is the automated anonymization of image material to avoid legal regulations concerning storage and use. This intention also motivates approaches covering the detection and blurring of vehicle registration plates, such as [9]. While data protection is not the prime objective of our research, the system we propose has no principal restriction, e.g. regarding runtime or reliability, which would preclude it.

Another purpose of image substitution is the generation of training data, in order to strengthen underrepresented classes or introduce special cases to existing datasets. Spata et al. [10] and Horn et al. [11] demonstrated the usefulness of automatically generated image data for training purposes w.r.t. the task of traffic sign recognition. Their CycleGAN-based approach has no need for detailed annotations but takes advantage of simple geometric relations that are immanent in traffic sign images, i.e., a planar foreground and a rather large distance to any background.

Generating license plate image has been covered in recent research as well. Liu et al. [12] present a general-purpose image generation system that learns separate representations for the image content, e.g., the license plate number, and its style. They provide results on several image datasets, among these Chinese license plate images. However, additional labelling, the plate number, is required for each image that is used in generator training. Wang et al. [13] demonstrate training a license plate classifier with generated images only. These are obtained via a style-transfer from artificially rendered Chinese license plates. As the authors use the style-transfer in one direction only, they describe several measures against mode collapse.

In the proposed method, we avoid relying on additional annotations. Instead we use a hybrid approach to: 1) incorporate formal knowledge, like the character patterns that form a valid license plate, 2) use well-established image processing routines to derive the key geometrical properties like scale and orientation of the license plate, and 3) embed the generated number plate into a real image inheriting its key characteristics and preserving the variance of the target dataset.

III. METHOD

The method presented in this paper is an extension to the approach by Horn and Houben [11], which uses a CycleGAN to perform a bidirectional style transfer on traffic sign images between cartoon and life-like domain. Their key idea is the transfer of real-life images into the cartoon domain, in which content can be easily replaced, such that a manipulated image transferred back to the life-like domain results in a realistic looking image with altered content. The cycle consistency inherent to CycleGAN training enforces the model to encode style information of real-life images such as background, illumination, and texture into barely visible intensity perturbations in the corresponding cartoon images while preserving image structures. The uniform background color encodes the appearance of the image background during CycleGAN training aiming to separate the generation of the license plate and its surrounding. The method has been adopted in this work to address the problem of exchanging license plates in camera images. The CycleGAN is further extended to additionally output a segmentation mask, which separates the license plate foreground from the image background. This mask is required to precisely reinsert only the license plate into the original image, and in comparison has been performed as an additional segmentation step in the original pipeline by Horn and Houben [11].

A. Generation Pipeline Overview

The proposed pipeline for exchanging license plates in camera images is illustrated in Fig. 2. Initially, a pre-trained object detector [14] is used to localize license plates in images of vehicles in arbitrary poses. Each detected license plate is cropped and resized to match the input resolution of the CycleGAN, i.e., 256×256 pixels. The license plate crop is propagated through the cartoon generator of the CycleGAN, which outputs a cartoonized version of the
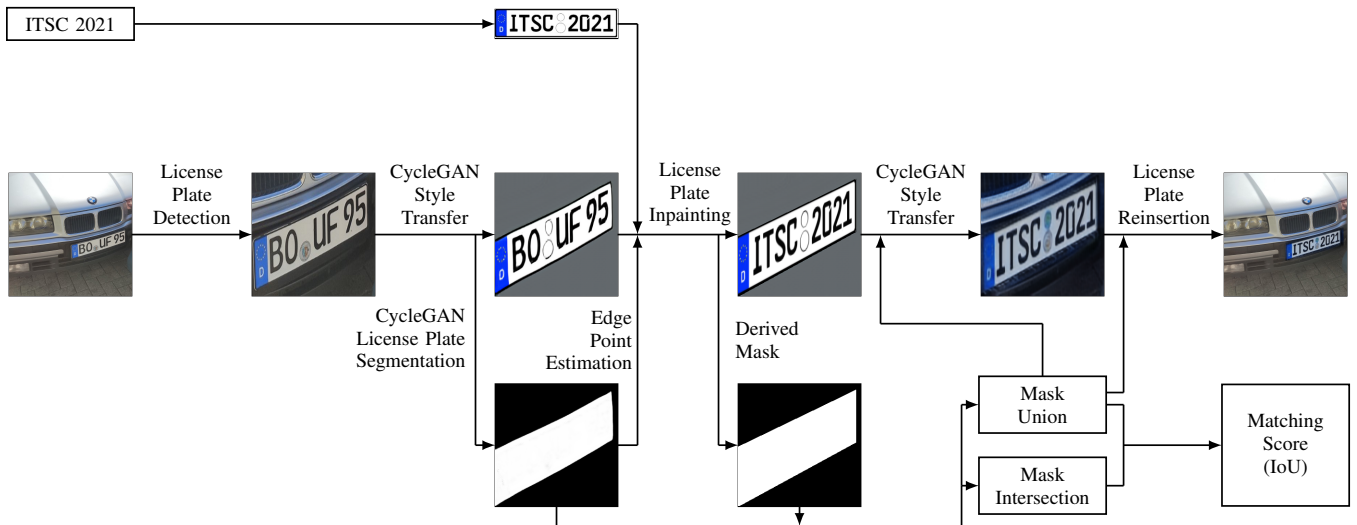
Fig. 2. Pipeline overview: For a given input image a license plate is detected, cropped and fed into the CycleGAN, which outputs a cartoon version of the input image as well as a binary license plate segmentation mask. The mask is used to determine the corners of the license plate, which are then adopted to render a new license plate template with a customized string, such as "ITSC 2021", in the same orientation on top of the cartoonized image. At the same time a second mask is derived from the detected edges, which is united with the original segmentation mask to achieve a more stable reinsertion process. The modified cartoon and mask are fed into the CycleGAN and the result is smoothly reinserted into the original image. Finally, a matching score between the two binary masks is calculated to serve as additional quality measure.

crop and an additional segmentation mask that separates the license plate foreground from the vehicle background.

Simultaneously, an alternative license plate number (e.g. "ITSC 2021") is rendered on a planar 2D template for German license plates using the official German license plate font. The template consists of a blue European registration plate with the letter D for Germany on the left-hand side and circular placeholders for the vehicle inspection and region identification stickers in the central part of the template. Notice that the license plate template and font varies across different countries and therefore requires localization for international use.

Based on the segmentation mask the four corners of the license plate are estimated by applying the Douglas-Peucker algorithm [15] on the mask's foreground. The algorithm extracts the contours of the license plate which are then propagated to a Shi-Tomasi corner detection [16]. The estimated corners are used to homographically project the customized license plate template onto the corners of the original one. The projected and filled template is then rendered on top of the initially cartoonized license plate crop and the segmentation mask for the template is generated.

Notice that we also tried to estimate the full 8-DOF perspective transformation of the license plate, which should lead to more accurate results but turned out to be unstable if estimated from a single image.

The cartoonized crop with the inpainted license plate template and the union of the original and the derived template segmentation masks are transferred back to the real-life image domain using the corresponding CycleGAN. The union of the segmentations ensures that the resulting segmentation mask covers the entire license plate content even if the template and the original license plate do not

fully align.

Finally, the segmentation mask is filtered with a Gaussian kernel and used as an alpha blending mask in order to smoothly reinsert the generated license plate into the original image. This form of gradual inpainting prevents edge artifacts that can occur at the boundaries through contrast differences between the generated and the original image.

Additionally, a matching score is calculated as intersection over union between the segmentation masks of the original and the template license plate. It quantifies how well the original and the template license plate are aligned in size, position, and perspective and therefore serves as an indicator for the success of the license plate substitution. We manually investigated matching scores with corresponding generated images and observed that scores above 0.85 usually reflect a sufficient alignment of the two masks. Approximately 68 % of the training data match or exceed this score. Following a traffic light scheme, we rated their confidence level as green. A matching score between 0.85 and 0.75 still leads to good results in many cases, but there is an increase of mismatched perspectives of inpainted license plate and background, which simplifies the identification of these images as generated ones. A confidence of yellow is assigned to those images, which covers approximately 18 % of the data. Finally, approximately 14 % of the data has a matching score below 0.75, which usually leads to a wrong perspective of the generated license plate and thus to inaccurate substitutions, so that a red confidence is given to those images.

B. Data Preparation and Model Training

The CycleGAN training requires two datasets, one with camera images of license plate close-ups and the other one with similar cartoon license plates. In real-world images, the

license plates were automatically cropped using the same pre-trained object detector [14] as used to localize the license plates in the proposed pipeline. This resulted in a total of 29,000 license plate close-ups, which vary in resolution, sharpness, contrast, lighting conditions, size, and perspective. Note that, as the license plate number is protected by data privacy regulations, the dataset cannot be made publicly available, but some examples with already replaced license plate numbers are shown in Fig. 1.

To generate the second dataset, the size, position, and orientation of all previously captured license plate close-ups were determined by estimating the homography between each license plate close-up and the corresponding generated image of the planar license plate template, which showed the same license plate number as the camera image. All successful estimations were stored and the license plate template subsequently used to render artificial cartoon license plate close-ups in perspective by randomly choosing one of the estimated homographies as well as an artificial, yet valid, German license plate number. Note that the latter is composed of a city code with up to three letters including umlauts, followed by one or two arbitrary letters, and a maximum of four digits. In order to generate the license plate cartoons, the values are sampled uniformly under the restriction that the final license plate number consists of at least four characters and eight characters at the most. The background color of the generated images is adapted to the mean intensity of a randomly chosen camera license plate close-up, which often results in a gray cartoon background. This approach ensured that the generated cartoon data features the same variations in size, position, and perspective as the real-life image dataset, but at the same time samples the space of possible license plate patterns more uniformly.

A main change to the CycleGAN implementation as originally proposed by Zhu et al. [17] is the removal of the identity loss, which is owed to the fact that the loss expects the same number of channels for the generator input and output – a prerequisite that is no longer matched as a result of the additional segmentation masks. Following the original publication, a batch size of 1 in combination with instance normalization has been used. The model was trained on the entire dataset with $29,000$ cartoon and $29,000$ real-world images using the ADAM optimizer ($\beta_1 = 0.5$, $\beta_2 = 0.999$) with a learning rate of 0.0002 for 23 epochs. The stopping criteria was determined by calculating the Fréchet Inception Distance [18] between the set of all generated images and the set of all training images. The score dropped from initially 252.3 with random weights to 21.8 after five epochs, 21.4 after 15 epochs, and 19.5 after 25 epochs. The lowest score of 17.4 have been reached after 23 epochs.

IV. Experiments

The proposed pipeline can generate an arbitrarily large amount of realistic license plate images with custom license plate numbers, which can be used for data augmentation, e.g., when training systems to recognize license plates numbers. If only a small amount of manually labeled training data

is available, we expect such a system to reach much higher recognition rates with additional artificial image data. To test this assumption we have trained artificial neural networks to recognize license plate numbers either on real-life images, generated images, or a combination of both. Moreover, at least for a small amount of vehicle images in which the license plate has been exchanged, humans should not be able to recognize manipulated images. Therefore, a user acceptance study has been performed in which humans are asked to decide whether an image with a possibly exchanged license plate number is manipulated or real.

A. Automatic License Plate Recognition

For supervised training of the artificial neural networks, a total of 45,000 real-world images with German licence plate strings as labels are available. In the dataset, the same license plate might appear more than once with shots taken from different angles, so that the number of unique license plates in the dataset amounts to $22,000$. $5,000$ images with different licence plate numbers have been randomly selected as an independent test set. From the remaining data all images showing the same license plate number as the one selected for the test data and images with a pipeline confidence level of red have been removed, so that in total 30,090 images remained for training. For augmentation we generated 60,000 images, processing each real-life image twice with random but valid license plate numbers, using the proposed pipeline on the entire training data and additional unlabeled data, in which only generated images with confidence green and yellow have been selected.

The network for license plate recognition is based on the pre-trained Xception architecture [19], from which the last convolution layer is selected and a mean average pooling layer followed by a fully connected layer is added. The output consists of eight softmax layers – one for each license plate character position. Each softmax layer is 39-dimensional, in which each dimension is representing one of the characters [A-ZÜÖ0-9] or a white space.

Several experiments are conducted to compare the performance on the $5,000$ test images with unique license plate numbers. Three sets of experiments cover training subsets of real-life images, generated images, as well as mixed data origins. For all experiments we measured the total categorical accuracy, i.e., the accuracy of recognizing all characters without mistake. Considering the possible variance immanent to the training process, displayed results refer to the mean of ten training routines each.

B. User Acceptance Study

The acceptance study was performed with 48 participants in total, each of whom was shown 50 real-life images featuring vehicles in perspective view with their original license plates and 50 manipulated images in which the license plates had been substituted. The real-life images were selected randomly from the previously defined test data, with the restriction that images were rejected manually beforehand if they were obviously outside of the desired input domain, i.e.,

grayscale images, images under extreme lighting conditions, or heavily blurred images. In these cases inpainting would have easily been identified by the participants, facilitating their decision process by elimination. This preselection concerned two percent of the images.

For the generated images, only those with a green matching confidence were selected. In these cases the pose estimation of the license plates worked rather well, reducing the chance of artifacts that occur when the template does not fully hide the original license plate. All images were resized such that the height equals 512 pixels as it would otherwise be possible to identify a manipulated image only based on the resolution difference between high resolution original images and the inpainted generated license plate crops, which have a native resolution of 256×256 pixels.

The 100 test images were shown in the same random order to each participant, who had 1.5 seconds to observe each image and decide afterwards whether the current image was real or generated. Similar to [2], a time restriction was introduced as – given enough time – it is almost always possible to find small artifacts or inconsistencies in images that are artificially generated by GANs. This is the case even if the model is significantly more complex and relies on more training data such as the StyleGAN[1] [20], for example. We therefore decided to limit the time given to force participants to rely on the overall visual impression rather than scene or expert knowledge, e.g., that the vehicle inspection sticker is always on the rear license plate in Germany – a piece of information not present in the used training dataset. The time restriction also prevents the participants from zooming into the image to actively search for those artifacts. Furthermore, to prevent people from identifying a generated license plate just by an unlikely configuration of letters, the city code was sampled from the training data and random alphanumeric codes were added such that the total length of the license plate string varied in a range of 6-8 characters.

V. RESULTS

A detailed discussion on both experiment types is given in the following. While the license plate recognition results mainly support findings from other research in this field, e.g. [11], the user acceptance study provides a differentiated view of the human perception of generated vehicle registration plates.

A. Automatic License Plate Recognition

The results of the automatic license plate recognition experiments are summarized in Fig. 3, in which a gradual increase of the training set size accompanied by a growing test accuracy can be observed. The curves show the average total accuracy on $5,000$ real test images and the corresponding standard deviation, averaged over 10 trials. The networks were trained on either real, generated, or a balanced combination of both image sets.
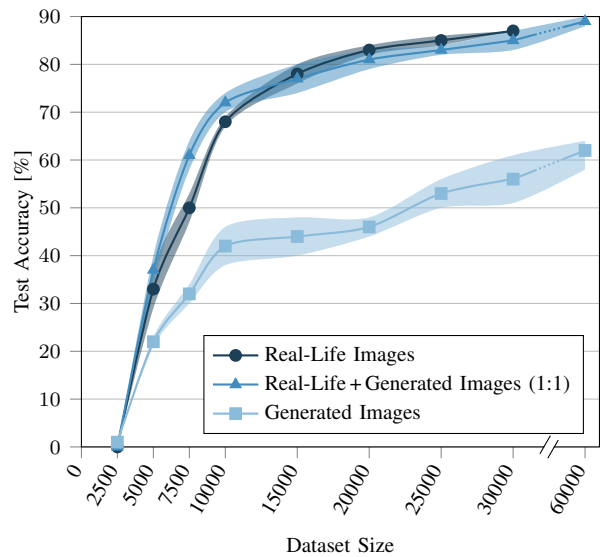
Fig. 3. Evolution of the total accuracy, i.e., the proportion of license plates for which every single character was correctly predicted. Accuracy and standard deviation for each depicted data point are based on 10 trials. The models were trained on either real-life, generated, or a balanced combination of real-life and generated images and evaluated on the same test set consisting of $5,000$ real-life images.

The figure clearly illustrates that using a training set size of $2,500$ or less is not sufficient for any of the networks. Those trained on $5,000$ images, on the other hand, already achieve accuracies of 22%, 33%, and 37% for generated, real and a combination of both types, respectively. When the generated images are used in combination with real ones for training, the performance is mostly comparable to the networks trained entirely on real images. However, for less than $15,000$ samples a combination of real-life and generated images is more favorable to the resulting accuracy than a pure real-life sample approach. This advantage cannot be maintained for bigger datasets, although accuracies of real-life and mixed training data differ only slightly for the same training set size. When comparing both curves with respect to the amount of real-life images within the datasets, e.g., comparing $30,000$ mixed images with $15,000$ real-world samples, augmentation of the training data with generated images leads to a significantly better performance. This trend is continued, e.g., when $30,000$ real-life images are augmented with another $30,000$ generated ones leading to a performance increase of approximately 3%.

In comparison, networks trained only on generated images achieve significantly lower total test accuracies. This indicates that the generated images, although beneficial in augmentation, do not fully cover the distribution of real-life images. A possible reason is the fact that only generated images with a matching score of at least 0.85 are selected for training data generation, which excludes difficult perspectives and lighting conditions. However, those challenging images are still present in the real-world training samples and are also part of the test set, so that the network requires at least some amount of real-life images in training to be able
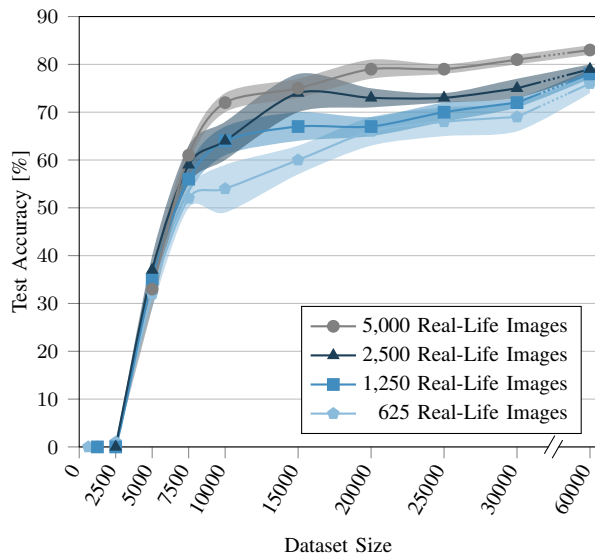
Fig. 4. Evolution of the total accuracy, i.e., the proportion of license plates for which every single character was correctly predicted. Accuracy and standard deviation for each depicted data point are based on 10 trials. Training datasets have a fixed amount of 625, $1,250$, $2,500$, and $5,000$ real-life images, respectively, and are expanded with a variable number of generated samples.
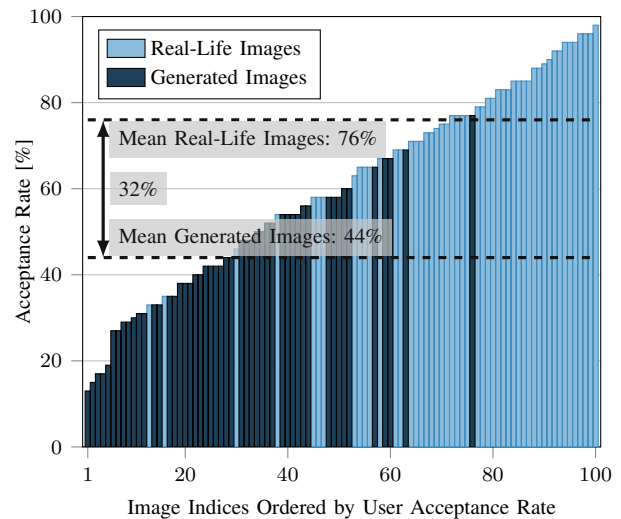

Fig. 5. Comparison of all test images ordered by their user acceptance rate, which is the proportion of participants that accepted an image as real whether it is real (light blue) or generated (dark blue).

to deal with those difficult samples. A clear advantage of using generated images in training is the additional variety in character combinations and thus increased robustness of the resulting classifier for rare cases.

To further investigate the effect of limited real-world data in training, multiple experiments on either 625, $1,250$, $2,500$ or $5,000$ real-life images in combination with a variable amount of generated samples have been performed. The results are summarized in Fig. 4, showing that all four variants perform better when the training set size is increased by additionally generated data. A higher amount of real-world images stabilizes the distribution w.r.t. to the test set and thus leads to higher accuracies, as well. With a relatively small effort of labeling only 625 images an accuracy of $76\,\%$ is already achievable, which is $14\,\%$ higher than training merely on generated images, and which further increases to $83\,\%$ when labeling $5,000$ images. This tendency can still be observed for vast amounts of real-world samples, as depicted in Fig. 3. Even for $30,000$ real-life images, the addition of generated data further improves the test performance to $89\,\%$ illustrating the value of the artificial data generated with the pipeline.

B. User Acceptance Study

The previous experiments have shown that generated images with a matching score above 0.85 are beneficial for data augmentation when training neural networks. However, this does not necessarily mean that those generated images are realistic from a human perspective, which in fact is not even required since neural networks are rather robust against small image artifacts, like those possibly inserted by the GAN. The human visual perception, in comparison, is very sensitive to

small inconsistencies even if the overall impression of the image is considered to be realistic. The user acceptance study has therefore been performed to estimate the realism of the generated images and investigate how often generated images are accepted as real.

Fig. 5 shows the average acceptance rate, i.e., the percentage of participants accepting a given image as real, regardless of whether it is actually real (light blue) or generated (dark blue) in increasing order. On average, $76\,\%$ of the real-life images are correctly identified as real, while $44\,\%$ of the generated images have been classified as real, as well. Thus, there is a gap of $32\,\%$ in order to reach the Nash equilibrium, in which participants cannot differentiate between real and generated images anymore. The difference in performance for real and generated samples can also be identified from the location of the respective samples on the abscissa. Most of the generated images have a lower score and are located toward the left-hand side, while the real-world ones are located with a higher score toward the right-hand side of the diagram. In case of the Nash equilibrium, real and generated images would be distributed uniformly across the acceptance rates. It is noteworthy that on average $24\,\%$ of the real images – in the presence of generated images – are falsely classified as generated. This gives evidence that the decision might be rather subjective than objective in a number of cases, i.e., not based on inconsistencies arising from the generation process but on the overall impression of an image.

Some of the generated images appear to be more realistic than others as their acceptance rates vary significantly from the average. In an extreme case, a generated image has been classified as real by $77\,\%$ of all participants, which, after manual investigation, appeared to be very realistic to the authors, as well. On the other hand, the two real-life outliers with $32\,\%$ and $34\,\%$ acceptance rate appeared to be rather unrealistic to the authors due to unusual lighting conditions,
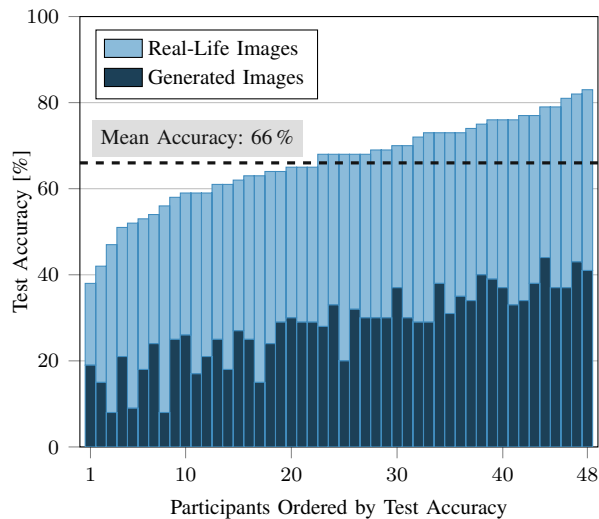
Fig. 6. Comparison of participants' total test accuracy, which is composed of the accuracy that real-life images are correctly classified as real (light blue) and the accuracy that generated images are correctly classified as generated (dark blue).

which might explain why most of the participants classified these images as generated. It also shows that the decision of whether an image is generated or not does significantly depend on the conditions a photo has been taken in.

Besides the varying results of the individual images, the performance of the participants differs due to their level of experience with generated data. The ascending order of the average test accuracy of each participant is shown in Fig. 6, which also illustrates the cumulated results of correctly classified real and generated images. On average a test accuracy of 66 % is achieved, with a peak performance of 84 % and a minimum of 38 %, showing indeed a substantial difference in performance across the participants. Furthermore, none of the participants could identify all generated images or all real-life images.

In order to further understand the difference in performance, some of the best and worst performing participants have been asked, which features they were focusing on for their decisions. While participants with a good performance clearly named artifacts produced by the GAN, such as missing details on the European registration plate, wrong coloring of the vehicle inspection sticker, or blurry edges caused by the inpainting, participants with a low performance mainly focused on the overall plausibility of the image such as unreasonable contrast variations. This further emphasizes our assumption that human beings are good in identifying small inconsistencies in images generated by GANs, which, however, does not necessarily mean that the images do not appear realistic as a whole. Nevertheless, there is still room for improvement, especially with regards to the level of detail during generation in our proposed pipeline.

VI. CONCLUSION AND FUTURE WORK

We have demonstrated that the manipulation of image data in the cartoon domain is a viable way to guide the results of the proposed image generation process. At the same time, a careful embedding of the altered image crop into the original image enables the user to adopt key characteristics from an individual image and, by extension, sources of variance from the entire underlying dataset. User acceptance studies have indicated a high level of realism while experiments with the recognition system showed that the global distribution of features necessary for training have been retained.

Some rare failure modes, e.g., inpainting with misestimated plate poses, as well as a more accurate but also much more challenging pose estimate based on more than the four license plate corners, remain to be addressed. However, we are positive that the system will reach a reliability that renders it close to full automation. This and the fact that the approach shows potential for real-time applicability open up the possibility for further areas of applications: Real-time license plate anonymization would enable the storage and widespread use of traffic cameras whose recordings are otherwise protected by data protection regulations in many countries worldwide. Being able to openly provide or access these data sources, possibly live, would mean a boost in traffic safety, intelligent transportation research, and smart city management as a whole.

REFERENCES

[1] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning From Simulated and Unsupervised Images through Adversarial Training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2242–2251.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[3] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8798–8807.

[4] H. Abu Alhaija, S. Mustikovela, A. Geiger, and C. Rother, "Geometric Image Synthesis," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2018, pp. 85–100.

[5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 740–755.

[6] Z. Ren, Y. J. Lee, and M. S. Ryoo, "Learning to Anonymize Faces for Privacy Preserving Action Detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 620–636.

[7] Ö. Sümer, P. Gerjets, U. Trautwein, and E. Kasneci, "Automated Anonymisation of Visual and Audio Data in Classroom Studies," *arXiv:2001.05080 [cs.CV]*, 2020.

[8] H. Hukkelås, R. Mester, and F. Lindseth, "DeepPrivacy: A Generative Adversarial Network for Face Anonymization," in *Proceedings of the International Symposium on Visual Computing (ISVC)*, 2019, pp. 565–578.

[9] L. Y. Chan, A. Zimmer, J. Lopes da Silva, and T. Brandmeier, "European Union Dataset and Annotation Tool for Real Time Automatic License Plate Detection and Blurring," in *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC)*, 2020, pp. 2220–2225.

[10] D. Spata, D. Horn, and S. Houben, "Generation of Natural Traffic Sign Images Using Domain Translation with Cycle-Consistent Generative Adversarial Networks," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 622–628.

[11] D. Horn and S. Houben, "Fully Automated Traffic Sign Substitution in Real-World Images for Large-Scale Data Augmentation," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 194–200.

[12] S. Liu, C. Zhao, Y. Gao, J. Wang, and M. Tang, "Adversarial Image Generation by Combining Content and Style," *Image Processing*, vol. 13, no. 14, pp. 2716–2723, 2019.

[13] X. Wang, Z. Man, M. You, and C. Shen, "Adversarial Generation of Training Examples: Applications to Moving Vehicle License Plate Recognition," *arXiv preprint arXiv:1707.03124*, 2017.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[15] D. H. Douglas and T. K. Peucker, "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.

[16] J. Shi and C. Tomasi, "Good Features to Track," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.

[17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[18] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017, vol. 30, pp. 6626–6637.

[19] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.

[20] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4401–4410.