Enhancing Object Detection Robustness for Cross-Depiction Through Neural Style Transfer

Francesca Fiani[1], Adriano Puglisi[1] and Christian Napoli[1,2,3]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, 00185 Roma, Italy
[2]Institute for Systems Analysis and Computer Science, Italian National Research Council, 00185 Roma, Italy
[3]Department of Computational Intelligence, Czestochowa University of Technology, 42-201 Czestochowa, Poland

Abstract

Modern neural networks models for computer vision are trained on millions of images. The idea is that models are able to increase generalization when the dataset contains well diversified images, e.g. with varied illumination and environmental conditions of the same objects. Generalization is particularly relevant in object detection, especially for what concerns the cross-depiction problem. In this work we explore the use of Neural Style Transfer as a novel technique to morph the original data, with the aim to enhance model generalization. To verify the effect on performances for object detection models, we selected the Faster R-CNN model to be applied on the Pascal VOC 2012 dataset. A number of tests were performed through style variations on images and by tuning Neural Style Transfer parameters to maintain the content of the original images. The experiments showed promising results, which effectively provide a foundation for future studies on cross-depiction via Neural Style Transfer.

Keywords

Neural Style Transfer, Object detection, Faster R-CNN, Pascal VOC 2012

1. Introduction

Object detection is a challenging task in computer vision which has a wide range of possible real-life applications, ranging from autonomous driving and healthcare to entertainment [1, 2]. This problem, while relatively new, has already been tackled in literature with several different approaches [3, 4]. The solutions are mainly classifiable in conventional methods, which are comprised of three phases (region selection, feature extraction and classification), and deep learning based methods [5]. The most advanced approaches focus on the use of deep neural networks, in particular convolutional neural networks (CNN), with the most popular solution to object detection being YOLO [6], developed in the years up to YOLOv8 [7]. Achieving high performance in this task is fundamental for several applications, with some examples being forensics or real-time usage (e.g. for autonomous driving). In order to improve the effectiveness of object detection models, various solutions to enhance generalization in unforeseen situations have been developed, the main ones being data augmentation and Neural Style Transfer. Data augmentation encompasses many different basic techniques, such as linear transformations, rotations and flipping, random cropping, random noise and brightness modulation. By applying these transformations to the original images, the data augmentation process generates new training data, therefore increasing the initial training data's variability and diversity to improve response to unseen images. One common challenge in object detection is dealing with noisy images. These are images that contain various types of distortions, such as blurring, noise and compression artifacts. Data augmentation can mitigate the effects of these distortions by generating new images with such features, thus making the model more robust to noisy inputs. Despite the success of these methods, however, accurately localizing small objects or objects with complex shapes, as well as dealing with occlusions and cluttered backgrounds, still present a challenge. Moreover, as proved by adversarial attacks, even state-of-the-art models can very easily miss the recognition of an object with basic manipulation on part of the image [8]. For this reason, different data augmentation techniques have been developed to face the aforementioned issues. Neural Style Transfer is one such solution and one of the most popular ones. Style transfer consists of the ability of models to transfer the style of one image to another. Before the advent of neural networks, style transfer applications were realized through several traditional methods such as region-based techniques, stroke-based rendering, example-based rendering and image processing and filtering [9, 10]. Such methods originally aimed at non-photorealistic rendering, and only later shifted towards the artistic stylization of 2D images, which is the pivotal concept on which Neural Style Transfer is built on. This process has been called image-based artistic rendering (IB-AR) [11]. Modern Neural Style Transfer, instead, makes use of two different starting im-

ICYRIME 2023: 8th International Conference of Yearly Reports on Informatics, Mathematics, and Engineering. Naples, July 28-31, 2023
fiani@diag.uniroma1.it (F. Fiani); puglisi@diag.uniroma1.it (A. Puglisi); cnapoli@diag.uniroma1.it (C. Napoli)
0009-0005-0396-7019 (F. Fiani); 0000-0002-9421-8566 (C. Napoli)

ages, the content image and the style image. The first one defines the context on which the artistic style depicted in the second must be applied, thus generating a new hybrid image with the content of the content image but the style of the style image. The generation is performed by a CNN, with the first tests being performed with a classic VGG19 model [12, 13]. From then, a whole taxonomy of algorithms for Neural Style Transfer was developed, broadly divisible in Image-Optimization-Based Online Neural Methods, which rely on multiple executions of the image optimization and gradual online style transfer, and Model-Optimization-Based Offline Neural Methods, which perform a single forward pass after optimizing the model offline. Starting from Neural Style Transfer, several sub-applications were derived. Some examples are Visual Style Modeling, which aims at synthesizing textures from images, and Image Reconstruction, which instead tries to reconstruct whole images from extracted fragments. This paper, instead, tackles a particular case of interdisciplinary task between object detection and Neural Style Transfer referred to as the cross-depiction problem [14]. Cross-depiction consists of recognising visual objects regardless of their form and style, and it's still an under-researched problem in computer vision. This translates to the capability of a neural network to correctly identify objects portrayed in artistic styles that are more or less different from their realistic representation in photographs. A neural network trained in the usual way will struggle to recognize a dog painted in an abstract way. To perform cross-depiction, the network will have to consider less specific features and focus on the shape of the dog itself, as well as other features that are not necessarily typical of realistic photos. Our aim is to train and fine-tune our object classification model to be more focused on the shape of the objects and on more generic features that would not be considered, or would be considered with a minor weight, in a conventional environment. In this work we show how applying style transfer on a particular dataset with different hyperparameters can increase the performance of a model like Faster R-CNN on a object detection classification task. By augmenting the data already present in the dataset with Neural Style Transfer transformations, the model is made more robust to outliers and edge cases, therefore rendering it applicable to more general situations. In particular, we focus on the application of Faster R-CNN on the Pascal VOC 2012 dataset, performing different tests to verify the preservation or improvement of the performances of the model after the application of Neural Style Transfer on the dataset. A subset of the total images was chosen to apply Neural Style Transfer on and be used as a test set. The experiments show that the style variation during training positively affects the performances of object recognition on a dataset of artistic images, cementing our approach as a possible solution to the cross-depiction problem.

2. Related Works

Several data augmentation techniques have been presented in modern deep learning as an efficient solution to improve model performances and limit overfitting during training [15]. Models, however, require substantial amounts of data in order to learn to classify images correctly, and the inability to provide this data usually correlates with poor performances during inference. The idea of using Neural Style Transfer as a form of data augmentation is not new, and it has already been verified as a domain-agnostic approach, making it suitable for various image classification tasks with several models (ResNet, VGG19 and Inception) [16]. One of the main problems in the original paper on Neural Style Transfer was the time needed by the algorithm to apply the style transfer, among the longest in all available Neural Style Transfer approaches [12, 9]. Following papers therefore showed how to increase the speed at which the style transfer is applied to the original image using a feed-forward approach, reducing the strain on the resources available for training purposes [17]. However, this method is only able to reproduce one style per model, and new, more flexible models were proposed to solve both problems. The category of Arbitrary-Style-Per-Model algorithms (ASPM MOB-NST) efficiently solves the scalability problem, with also the possibility of completely removing learning limitations through feature transform [18], but introducing less impressive results compared to more specific approaches [19, 20, 21, 18]. It has also been verified that Neural Style Transfer can be used to reduce bias, and a novel pipeline for Antibody Mediated Rejection classification has provided an implementation faster than current SOTA approaches [22]. One of the most robust choice for object detection is R-CNN, or Region-based Convolutional Neural Networks, which marked a significant breakthrough in object detection performance, outperforming many rival algorithms [23]. The key concept behind Region-based Convolutional Neural Network architectures is region proposals (RPNs), regions in the image that could contain an object of interest, which are then fed to a Convolutional Neural Network, typically a ResNet or a VGG. The extracted features are finally passed to a series of fully connected layers for the final predictions of the classification and the object detection. The largest drawback and bottleneck of the original R-CNN architecture is its computational expensiveness, as it requires running the CNN separately for each object proposal. Moreover, the selective search algorithm is fixed, which means that no learning happens at that stage. A whole family of state-of-the-art models spanned from R-CNN to address these issues, with architectures such as Fast R-CNN [24] and Faster R-CNN [25] building upon the previous model's success to improve object detection accuracy and speed. These models replace the separate CNN for each proposal with a shared CNN used to extract features for all the proposals, allowing faster processing. Also, instead of feeding the region proposals to the CNN, the same CNN generates both object proposals and detection. The difference between Fast R-CNN and Faster R-CNN is that the latter, instead of using the slower selective search algorithm on the feature map to draw the region proposals, utilizes a separate network to get the region proposals, further reducing execution time. Models like Faster R-CNN are able to perform relatively well when presented with images that resemble the ones seen during training, showing the capability to generalize and opening to the possibility of being fine-tuned for custom datasets.

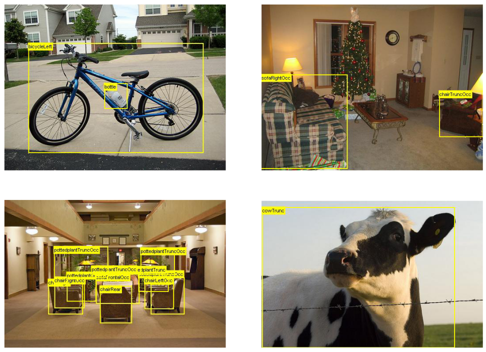

Figure 1: Examples of images from the Pascal VOC dataset. Each image has all the objects pertaining to the 20 object classes classified via bounding boxes, which are allowed to overlap. The labels are also indicative of the orientation of the object.

3. Implementation

Our work aims at presenting a novel approach and solution for the cross-depiction problem, with Faster R-CNN being a particularly good fit for our task. More precisely, the model that we used is the Faster R-CNN ResNet50 FPN from the Torchvision models, which combines the ResNet50 model as feature extraction backbone with a Feature Pyramid Network (FPN). This way, object detection performance is improved by generating a set of feature maps at different scales, which helps the model detect objects of varying sizes and aspect ratios. The experiments performed in our work are aimed at understanding how a CNN performs on unusual abstract images under various conditions, and how much it is able to generalize in the presence of non-realistic features, with the goal of achieving object detection in artwork-like images. This would present a solution to the cross-depiction problem by making an object identifiable regardless of the style of the image. To perform the task, we employed the Neural Style Transfer methods previously described to widely augment a well known dataset, Pascal VOC [26], used as a standard benchmark for evaluating object detection models. In particular, we used the 2012 version, the latest available. It contains 17,125 images annotated for object detection, as well as object classification and image segmentation. The images consist of 20 object classes, including animals, vehicles, and common household items. Some examples of images contained in the dataset are shown in Figure 1. A similar data augmentation has already been presented in previous works [27], but we won't focus solely on people recognition and the people class, instead employing the whole dataset.

Faster R-CNN ResNet50 FPN is deployed in its version pre-trained on ImageNet [28], a large-scale image database widely used in computer vision research, composed of 1000 classes. In order to be trained on the 20 Pascal VOC classes, the model is initialized replacing the last layer responsible for the regions of interest (RoI) with a new one that has 21 output features, the 20 classes plus the background. Before starting the fine-tuning, which is performed on the Pascal VOC dataset, a pre-processing phase in which the images are resized to the standard format of 256x256 pixels and the pixel values are normalized in the range [0, 1] was necessary. The dimensions of the bounding boxes' labels have also been adapted accordingly to keep the ratio with the resized images. After this, the network goes through the fine-tuning. This process is performed on 80% of the dataset, leaving the remaining 20% for evaluating it. The model uses stochastic gradient descent (SGD) as optimizer, with a starting learning rate of 0.001, a momentum of 0.9, and weight decay of 0.0005. The fine-tuning lasts 10 epochs. The learning rate first goes through a warm-up period of 1000 iterations in order to get to the starting learning rate in a gradual way. Then, it is adjusted over the fine-tuning period following a learning rate scheduler, with step size 4 and gamma 0.1, which are the values that best optimized the performances while also avoiding overfitting. We first created a variety of subsets of the initial Pascal VOC dataset through Neural Style Transfer. Specifically, 12 different artistic styles of different time periods were selected (e.g. Cubism or Puntinism), and the NST was applied in two versions, one with a lighter stylization and another with a stronger one. The intuition is that by performing the NST, one is able to produce a dataset of a desired style which is already labelled, since the position and the dimension of the bounding boxes of the objects remain unchanged. The parameters used to

Figure 2: Examples of images from the Pascal VOC dataset with NST applied on them. The images on the left are the style images and the images on top are the images from the Pascal VOC dataset. The combinations are the results of NST application.

obtain the enhanced datasets are total_ steps = 35 and learning_rate = 0.02 for the lighter stylization one and total_steps = 55 and learning_rate = 0.05 for the stronger stylization one, with alpha = 0.8 and beta = 0.3 in both versions. These values have been chosen as a compromise between a recognizable adaptation of the applied style and the preservation of the objects in the image, although in the stronger version objects of smaller dimensions are often distorted and unrecognizable. Some examples of results of NST application to the Pascal VOC dataset are shown in Figure 2. Afterwards, we performed several experiments to verify how our model acts when lighter or stronger stylized images are fed to it. We also study how it performs when trained in different ways on the previously produced subsets of stylized images, evaluating it both on the light and the strong stylization, and both on seen and unseen styles. To evaluate the results of the experiments, we used a group of average precision and recall measurements that can estimate the performance of the object detection at various levels of overlap between the predicted bounding boxes and ground truth ones. The standard metric is the mean average precision (mAP) with a 50% bounding box overlap with the labelled box [26]. The performance has been evaluated with average precision AP, AP50 and AP75 as they are defined in the COCO detection evaluation metrics. AP is the average precision value at different thresholds of intersection over union (IoU), respectively 0.50 for AP50, 0.75 for AP75 and 0.50 to 0.95 for AP, evaluated for maximum detection of 100% for all areas. Separate AP scores are also available for different area sizes, divided into small, medium, and large objects, to measure the model's ability to detect objects of different sizes. The model pre-trained and fine-tuned only on the original Pascal VOC performs very poorly, with an AP50 of 0.318 on the light stylization and with an AP50 0.147 on the strong one. With just 10 epochs of training on the stylized images, however, the evaluation gets to AP50 0.549 on the light stylization and 0.356 on the strong one, which is already a good result compared to similar experiments [29]. We found that the best way to train the model is to conduct the training with a group of subsets of stylized images and a group of normal photographs at the same time. This keeps the object recognition grounded to a certain degree of reality, reducing weight assignment to some features and maintaining a slightly better ability of generalization. The AP50 score with the mixed training set is 0.553, with respect to the 0.530 of the model trained only on the stylized images, and a better score over the original test set was also maintained. It is possible to achieve even better results by training for more epochs, but to avoid overfitting on the training images we will use the fine-tuned model weights with 10 epochs as a starting point. The next experiments aimed at evaluating the performances of the aforementioned models over other images of different styles. We trained the model on eight of the subsets, leaving the remaining four styles for the test, with subsets composed of images unseen during the training. The final results evaluated in AP50 are 0.525 as the average of the scores obtained with light NST, and 0.247 with strong NST. In both cases the model fine-tuned on the light stylization has been used. For the model fine-tuned on a training set of strong stylization, instead, we got an AP50 of 0.519 and 0.316 on light and strong NST respectively. These results show that correspondence between the fine-tuned models and the training set positively reflects on the performance of the object detection. In Table 1 we show an overview of the results for each subset obtained from the model fine-tuned on the strong stylization and tested on light stylization comparing the different metrics (AP, AP50 and AP75). It is possible to observe that after this type of fine-tuning the model obtains a certain degree of generalization, showing detection performances on the last four unknown styles which are in line with the results obtained for the other classes.

4. Conclusion

The analyzed results confirm the concrete possibility to achieve data augmentation on images with varied artistic styles for any given dataset. We also demonstrate that a CNN is able to generalize under the presence of different features derived from different styles, therefore confirming the effectiveness of this method. This opens up to several possible applications, such as performing

Table 1
Average precision of Faster R-CNN fine-tuned on strong NST applied on subsets of PascalVOC with light stylization.

Dataset Subset	AP	AP50	AP75
Cubism	0.383	0.690	0.380
Puntinism	0.262	0.527	0.218
Pop Art 1	0.332	0.621	0.304
Van Gogh	0.300	0.599	0.241
Yukhnovich	0.249	0.514	0.220
William Turner	0.164	0.383	0.122
Jackson Pollock	0.289	0.570	0.263
Futurism	0.291	0.564	0.286
Monet	0.267	0.581	0.185
Surrealism	0.207	0.415	0.172
Kandinski	0.346	0.607	0.342
Pop Art 2	0.238	0.509	0.169

mass object detection in large datasets such as museum collections or online databases of artworks (both amateur and professional), resulting in an automatic extraction of metadata related to the identification and localization of objects. This could also be extended to the automatic creation of new datasets of non-photographic images for object detection. In such a process, the analyzed technique can be employed as a starting point to outline the bounding boxes of the objects in the scene, which can be then verified and adjusted. Many fields of possible applications can derive from the ability of neural network models to accurately perform object detection in whatever form of non-realistic representation. By providing a model the possibility to track objects and monitor their behaviours in different environments with a coherent artistic style, one could apply it also to animation, videogames, etc. Another interesting application coming from the achievement of higher levels of abstraction in object detection is the capacity of future AI agents, like generative agents, to behave socially and simulate human patterns [30, 31, 32, 33],. This approach, if extended to textual tasks, can also allow agents to perform tasks which require the understanding of unlabelled and unseen representations of various types, for example to navigate online forums or any kind of website to perform data scraping in a more comprehensive way. With regards to possible future works, our biggest limitation was the lack of big annotated datasets of artwork images for object detection, and it would be likewise insightful to see the results of tests of a model trained on such images as well. Finally, it would be also useful to extend the style transfer method illustrated here to other computer vision tasks, such as image segmentation or pose estimation, and see how much of what has been commented also applies to these problems.

Acknowledgements

This work has been developed at is.Lab() Intelligent Systems Laboratory at the Department of Computer, Control, and Management Engineering, Sapienza University of Rome (https:// islab.diag.uniroma1.it). The work has also been partially supported from Italian Ministerial grant PRIN 2022 "ISIDE: Intelligent Systems for Infrastructural Diagnosis in smart-concretE", n. 2022S88WAY - CUP B53D2301318, and by the Age-It: Ageing Well in an ageing society project, task 9.4.1 work package 4 spoke 9, within topic 8 extended partnership 8, under the National Recovery and Resilience Plan (PNRR), Mission 4 Component 2 Investment 1.3—Call for tender No. 1557 of 11/10/2022 of Italian Ministry of University and Research funded by the European Union—NextGenerationEU, CUP B53C22004090006.

References

[1] V. Ponzi, S. Russo, V. Bianco, C. Napoli, A. Wajda, Psychoeducative social robots for an healthier lifestyle using artificial intelligence: a case-study, volume 3118, 2021, pp. 26 – 33.

[2] V. Marcotrigiano, G. D. Stingi, S. Fregnan, P. Magarelli, P. Pasquale, S. Russo, G. B. Orsi, M. T. Montagna, C. Napoli, C. Napoli, An integrated control plan in primary schools: Results of a field investigation on nutritional and hygienic features in the apulia region (southern italy), Nutrients 13 (2021). doi:10.3390/nu13093006.

[3] F. Bonanno, G. Capizzi, C. Napoli, Some remarks on the application of rnn and prnn for the charge-discharge simulation of advanced lithium-ions battery energy storage, 2012, pp. 941 – 945. doi:10.1109/SPEEDAM.2012.6264500.

[4] G. Capizzi, F. Bonanno, C. Napoli, A wavelet based prediction of wind and solar energy for long-term simulation of integrated generation systems, 2010, pp. 586 – 592. doi:10.1109/SPEEDAM.2010.5542259.

[5] R. Kaur, S. Singh, A comprehensive review of object detection with deep learning, Digital Signal Processing 132 (2023) 103812. URL: https://www.sciencedirect.com/science/article/pii/S1051200422004298. doi:https://doi.org/10.1016/j.dsp.2022.103812.

[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, arXiv (2015). arXiv:1506.02640.

[7] D. Reis, J. Kupec, J. Hong, A. Daoudi, Real-time flying object detection with yolov8, arXiv (2023). arXiv:2305.09972.

[8] H. Xu, Y. Ma, H. Liu, D. Deb, H. Liu, J. Tang,

A. K. Jain, Adversarial attacks and defenses in images, graphs and text: A review, arXiv (2019). arXiv:1909.08072.

[9] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song, Neural style transfer: A review, IEEE Transactions on Visualization and Computer Graphics 26 (2020) 3365–3385. doi:10.1109/TVCG.2019.2921336.

[10] F. Bonanno, G. Capizzi, G. L. Sciuto, C. Napoli, Wavelet recurrent neural network with semi-parametric input data preprocessing for micro-wind power forecasting in integrated generation systems, 2015, pp. 602 – 609. doi:10.1109/ICCEP.2015.7177554.

[11] J. E. Kyprianidis, J. Collomosse, T. Wang, T. Isenberg, State of the "art": A taxonomy of artistic stylization techniques for images and video, IEEE Transactions on Visualization and Computer Graphics 19 (2013) 866–885. doi:10.1109/TVCG.2012.160.

[12] L. A. Gatys, A. S. Ecker, M. Bethge, A neural algorithm of artistic style, arXiv (2015). arXiv:1508.06576.

[13] S. Pepe, S. Tedeschi, N. Brandizzi, S. Russo, L. Iocchi, C. Napoli, Human attention assessment using a machine learning approach with gan-based data augmentation technique trained using a custom dataset, OBM Neurobiology 6 (2022). doi:10.21926/obm.neurobiol.2204139.

[14] H. Cai, Q. Wu, T. Corradi, P. Hall, The cross-depiction problem: Computer vision algorithms for recognising objects in artwork and in photographs, 2015. arXiv:1505.00110.

[15] T. Kumar, A. Mileo, R. Brennan, M. Bendechache, Image data augmentation approaches: A comprehensive survey and future directions, 2023. arXiv:2301.02830.

[16] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Communications of the ACM 60 (2017) 84–90. doi:10.1145/3065386.

[17] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, arXiv (2016). arXiv:1603.08155.

[18] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, M.-H. Yang, Universal style transfer via feature transforms, arXiv (2017). arXiv:1705.08086.

[19] T. Q. Chen, M. Schmidt, Fast patch-based style transfer of arbitrary style, arXiv (2016). arXiv:1612.04337.

[20] X. Huang, S. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, arXiv (2017). arXiv:1703.06868.

[21] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, J. Shlens, Exploring the structure of a real-time, arbitrary neural artistic stylization network, arXiv (2017). arXiv:1705.06830.

[22] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, A. Smolic, Stada: Style transfer as data augmentation, in: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, SCITEPRESS - Science and Technology Publications, 2019. URL: http://dx.doi.org/10.5220/0007353401070114. doi:10.5220/0007353401070114.

[23] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, arXiv (2014). arXiv:1311.2524.

[24] R. Girshick, Fast r-cnn, arXiv (2015). arXiv:1504.08083.

[25] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, arXiv (2016). arXiv:1506.01497.

[26] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes (VOC) challenge, International Journal of Computer Vision 88 (2010) 303–338. URL: https://doi.org/10.1007/s11263-009-0275-4. doi:10.1007/s11263-009-0275-4.

[27] D. Kadish, S. Risi, A. S. Løvlie, Improving object detection in art images using only style transfer, arXiv (2021). arXiv:2102.06529.

[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

[29] N. Gonthier, S. Ladjal, Y. Gousseau, Multiple instance learning on deep features for weakly supervised object detection with extreme domain shifts, Computer Vision and Image Understanding 214 (2022) 103299. URL: http://dx.doi.org/10.1016/j.cviu.2021.103299. doi:10.1016/j.cviu.2021.103299.

[30] N. Brandizzi, S. Russo, G. Galati, C. Napoli, Addressing vehicle sharing through behavioral analysis: A solution to user clustering using recency-frequency-monetary and vehicle relocation based on neighborhood splits, Information (Switzerland) 13 (2022). doi:10.3390/info13110511.

[31] G. De Magistris, M. Romano, J. Starczewski, C. Napoli, A novel dwt-based encoder for human pose estimation, volume 3360, 2022, pp. 33 – 40.

[32] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, arXiv (2023). arXiv:2304.03442.

[33] G. De Magistris, R. Caprari, G. Castro, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Vision-based holistic scene understanding for context-aware human-robot interaction 13196 LNAI (2022) 310 – 325. doi:10.1007/978-3-031-08421-8_21.