Improving vision models with targeted synthetic data generation

Markiyan Kostiv

August 2024

Abstract

This PhD proposal explores the challenges and advancements in synthetic data generation for computer vision, particularly in narrow domains with limited data availability. The goal of my research is to enhance downstream task performance by incorporating domain-specific knowledge into large generative models, addressing biases and edge cases, and generating high-fidelity synthetic data to reduce reliance on manual labeling. The study will also investigate techniques to better align synthetic data with real-world distributions, thereby improving model robustness and fairness. The ultimate goal is to develop methods that produce more accurate and reliable models across diverse real-world applications.

1 Introduction

Over the last decade, the rise of the computer vision field has been heavily influenced by the success of utilizing large amounts of carefully curated human-labeled datasets. The diversity of the dataset, its distribution bias [1], and the correctness of human annotation [2], arguably, are crucial to creating reliable production deep learning models. However, this approach has limitations in the scarce scenarios and edge cases that are hard to capture in real-world data and in specific domains where data collection and annotation resources are limited, such as medical imagery, military, environmental monitoring, early disaster prevention, and others. To address these problems, the industry relies on creating high-quality simulations [3] [4] that are time-consuming and expensive to design and adapt to real-world distributions.

Synthetic data offers an alternative way of generating large volumes of diverse and high-quality vision data [5][6][7] that can cover a wide range of scenarios while simultaneously having precise labels that supervised deep learning models can efficiently utilize for training. The rise of generative adversarial networks and the recent success of large diffusion models have proven to be able to generalize well on large amounts of noisy data and utilize this knowledge in a controllable way [8][9].

However, challenges like a lack of domain-specific knowledge in narrow domains, potential bias in synthetic data, and domain gap between

generated and real-world scenarios remain [10]. Additionally, there is a variability issue in the data generated by the model that was trained on a limited distribution. Simultaneously, by increasing the data variance, it becomes challenging to control the data drift and keep the generated data on track with the real-world distribution [11]. In the following sections, I'll describe a possible way to mitigate those problems. In the Related Work section, I outline the existing approaches toward synthetic data generation and domain adaptation, covering the state-of-the-art approaches and research directions of academia and industry. In the Motivation section, I describe the importance of the problem and its broader social and economic impact. The subsequent Methodology section outlines my perspective on enhancing the current approaches toward better synthetic datasets by utilizing large amounts of widely available noisy data and transferring learned capabilities into narrow domains. Finally, in the Conclusion section, I summarize the key elements of this proposal and evaluate the various aspects of the potential impact of my research.

2 Related Work

There are various approaches to generating high-quality synthetic data, ranging from creating comprehensive scenes and world models with game development engines like Unreal Engine [12], applying domain adaptation techniques to reduce the data shift between synthetic and real-world domain [13], or utilize deep learning models, such as latent diffusion models [14] or generative adversarial models [15]. While designing realistic scenes by hand can provide high-quality data, it remains a very time-consuming and expensive approach. Therefore, it usually limits the amount of data that can be created this way. On the other hand, domain adaptation and data generation can drastically reduce the data collection cost and offer unlimited data points.

Generative models for vision-data synthesis. Conditional adversarial networks [16, 17, 18] were one of the first methods that achieved plausible perceptual quality with the high-resolution image synthesis in various domains.

It was later shown that latent diffusion models (LDMs) are superior to GANs in the image synthesis tasks [19]. By operating in the latent space, they reduce the computational demands and can learn from the large-scale datasets collected on the internet [20]. Besides using text embedding [21] for conditional generations, these models can generalize well to additional inputs with much smaller datasets and compute requirements with techniques like ControlNet[8], DreamBooth [9], and T2I [22]. Following the success of single image generation, this approach was extended to video generation [23, 24], and can provide high-quality generation while maintaining temporal consistency and effectively model both short and long-range dependencies.

Application of generative models in synthetic data. The use of synthetic data has become an industry standard for solving various computer vision tasks, such as human-pose estimation [25], text localization and recognition [26], multi-object detection and tracking for au-

tonomous driving [27]. Generative vision models can be used to synthesize datasets that are on par with human-labeled datasets. For example, the saliency detection model trained on purely synthetic data [28] achieves 98.4% F-measure compared to the model trained on the largest open-source saliency detection benchmark dataset DUTS-TR [29]. Such approach is proved to be successful in other downstream computer vision tasks, such as semantic segmentation [30], depth estimation [31], object detection [32], and classification [33].

It also was shown that these large models trained on billions of images could be fine-tuned and utilized for data generation in the domains that are data-scarce, such as medical imagery [34, 35]. Additionally, they outperform other approaches in image-to-image translation for domain adaptation tasks in data-scarce domains [36] and in one-shot unsupervised domain adaptation [37].

Challenges in learning from the synthetic data. Despite the recent success of generative models in synthetic data, there are significant challenges to training robust models on downstream tasks only using artificial data. It is shown that although downstream models trained on synthetic data generated by vision language foundation models (VLFMs) show impressive capabilities on the academia benchmarks, their capacity is significantly diminished in the real-world scenarios [11]. The primary reason is that academic benchmark datasets are akin to the data scraped from the Internet that was used to train VLFMs. Therefore, the question arises if the generators are capable of introducing any advantage over the training on the relevant upstream data and if generators simply interpolate the original distribution and do not enhance it [10]. To solve the challenge of the bias in the upstream data and extrapolate the distribution, the generators are manipulated to generate bias-conflict synthetic samples [38].

3 Motivation

Despite the recent advances in generative models and overall interest in synthetic data generation in the industry, there are still areas that require further research.

Narrow domains often lack extensive data availability found in the Internet-scale dataset and face significant challenges in fully leveraging supervised learning models [39]. The limited size of these datasets typically restricts the development of robust models capable of performing well in real-world scenarios [40]. The current advancements in the synthetic data and generative models research have yet to overcome this limitation, as the generated data points do not provide gain over the targeted sampling from the upstream pool they originate from. Rather than introducing a new variety, these synthetic data points tend to interpolate within the existing distribution, offering little improvements for downstream tasks [10]. Therefore, one of the main goals of my research is to explore methods of utilizing the knowledge obtained by the generator trained on the large Internet-scale datasets on the targeted domain and incorporate domain-specific knowledge into the generator to create more

comprehensive training data to boost downstream performance.

Traditional datasets often have biases that come from the uneven representation of various classes or scenarios, leading to models that perform well on common cases but struggle with rarer, more critical situations [41]. These biases are particularly problematic in domains where the cost of error is high, such as in medical diagnosis, autonomous driving, or military applications. Moreover, edge cases, rare but crucial instances, are frequently underrepresented in the data, further deepening the problem [42]. To address these issues, my research will focus on identifying such edge cases and biases in the target distribution and leveraging synthetic data generation to enhance the diversity of training datasets by introducing more edge cases and counteracting inherent biases. By creating synthetic data that specifically targets these underrepresented scenarios, the goal is to produce models that are not only more robust but also fairer and more accurate across a broader range of real-world applications.

Manual labeling of datasets, particularly in complex scenarios such as dense object detection [43], is a resource-intensive process, often requiring significant time and expertise. In tasks where objects are densely packed or where precision is critical, such as in medical imaging, autonomous driving, or aerial surveillance, the demand for accurate labeling becomes even more pronounced. The need for domain-specific knowledge further increases the cost. This not only drives up the financial costs but also limits the scale and speed at which high-quality labeled data can be produced. To address these challenges, my research will explore the use of synthetic data generation as a cost-effective alternative to manual labeling. By generating high-fidelity synthetic data replicating these complex scenarios, we can significantly reduce the reliance on manual labeling while maintaining, or even improving, the quality of the training data.

4 Approach

The methodology of this research will be structured to achieve three main objectives: enhance the performance of the downstream tasks in narrow domains using domain-specific datasets, address biases and edge cases inherent in the large real-world datasets, and generate high-fidelity synthetic data for challenging and costly labeling scenarios.

4.1 Enhancing Synthetic Data Generation for Narrow Domains

The first step in this research will be to refine the synthetic data generation process for narrow domains where data scarcity is a significant challenge. To achieve this, I will leverage large generative models, such as latent diffusion models pre-trained on Internet-scale datasets like LAION-5B [20], and fine-tune them with various adapters inspired by T2I-Adapter [22] and ControlNet [8] to incorporate domain-specific knowledge [44], thereby enhancing the original small distribution of the target domain. Next, I'll further explore how we can improve the existing steering techniques for

large generative models that can better utilize specific target-domain features. This approach will include image-to-image translation tasks for domain adaptation, text-to-image generation for a wider variety of distributions, and a number of ablation experiments with synthetic data pre-training and fine-tuning. I will assess the impact of such data on downstream tasks, such as object detection and semantic segmentation, by comparing model performance with and without the use of synthetic data to evaluate its impact on robustness. To further enhance the variety within the small target domain dataset, I will explore techniques to better align the generated data with the target distribution [45] and address its low variety by generating more bias-conflict examples [38].

4.2 Addressing Biases and Edge Cases in Synthetic Data Generation

The focus of this research direction will be on identifying and mitigating biases present in common real-world datasets and addressing edge cases that may compromise model performance. Biases in datasets can lead to models that excel in standard scenarios but underperform in less frequent yet critical situations. Similarly, edge cases—rare or particularly challenging instances—can significantly affect the robustness of the model in the real-world setup.

To tackle these challenges, I'll begin by analyzing real-world datasets for downstream tasks, such as COCO [46], CityScapes [47], DOTA[48], nuScenes [49], and similar, to identify the challenging examples for state-of-the-art downstream task models. To achieve that, I will use techniques to extract high-uncertainty and high-loss samples [50] [51] in both training and validation datasets. These instances are likely to represent either mistakes in annotations or scarce scenarios. I will further apply techniques inspired by diversity-based sampling [52] to find the candidates for a generation. After the evaluation of such examples, I will build a synthetic data generator similar to the approaches described in 4.1, enrich training data with more examples similar to the challenging scenarios, and measure the performance of the downstream models after data enrichment.

4.3 High-Fidelity Synthetic Data for Challenging Labeling Scenarios

The final objective of this research is to explore how the generation of high-fidelity synthetic data can reduce the efforts and error rate associated with manual labeling in challenging scenarios, such as dense object detection [43], image matting [53], remote sensing through semantic segmentation [54], and others. This research will investigate whether simulation techniques [55] when combined with domain adaptation methods [56], can effectively substitute the need for manual annotation. Furthermore, we will assess whether the approach described in 4.1 can further accelerate the acquisition of such data by generating high-fidelity samples that meet the specific requirements of these complex labeling tasks by using large generative vision models. We will explore how the data

enrichment by such examples impacts the performance on common downstream tasks and measure the performance on challenging datasets such as CrowdHuman [57], UCF-QNRF [58], DOTA [48], and similar.

5 Conclusion

In this PhD proposal, I highlighted the key challenges to the effective use of synthetic data for training supervised deep learning models and outlined my approach to advance the field. The research is focused on three primary objectives: enhancing the performance of downstream tasks in narrow domains, addressing biases and edge cases inherent in large real-world datasets, and generating high-fidelity synthetic data for scenarios where manual labeling is particularly challenging.

By improving the process of synthetic data generation and incorporating domain-specific knowledge, my approach aims to overcome the data scarcity limitations in narrow domains. Furthermore, by identifying and mitigating biases and edge cases, the proposed research seeks to improve the robustness and fairness of models in real-world applications. Finally, the development of high-quality synthetic data generation techniques promises to reduce the dependency on manual labeling, thereby accelerating the development and adoption of machine learning models. This advancement represents a major step towards the next-level machine learning democratization, enabling broader access to powerful AI tools and facilitating innovation across diverse industries.

References

[1] Tianda Wang et al. "Towards fairness in visual recognition: Effective strategies for bias mitigation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 8919–8928.

[2] Xiaoxiao Xu et al. "How Do Label Errors Affect Thin Crack Detection by DNNs". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2023, pp. 2307–2316.

[3] Ben Alvey et al. "Simulated photorealistic deep learning framework and workflows to accelerate computer vision development for autonomous vehicles". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2021, pp. 2331–2338. URL: https://openaccess.thecvf.com/content/ICCV2021W/WAAMI/papers/Alvey_Simulated_Photorealistic_Deep_Learning_Framework_and_Workflows_To_Accelerate_Computer_ICCVW_2021_paper.pdf.

[4] Yifan Li et al. "MatrixCity: A large-scale city dataset for city-scale neural rendering and urban scene understanding". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023, pp. 2917–2927. URL: https://openaccess.thecvf.com/content/ICCV2023/papers/Li_MatrixCity_A_Large-scale_City_Dataset_for_City-scale_Neural_Rendering_and_ICCV_2023_paper.pdf.

[5] Zhihang Song et al. "Synthetic Datasets for Autonomous Driving: A Survey". In: *IEEE Transactions on Intelligent Vehicles* 9.1 (Jan. 2024), 1847–1864. ISSN: 2379-8858. DOI: 10.1109/tiv.2023.3331024. URL: http://dx.doi.org/10.1109/TIV.2023.3331024.

[6] Jonathan Tremblay, Thang To, and Stan Birchfield. *Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation*. 2018. arXiv: 1804.06534 [cs.CV]. URL: https://arxiv.org/abs/1804.06534.

[7] Samarth Mishra et al. *Task2Sim : Towards Effective Pre-training and Transfer from Synthetic Data*. 2022. arXiv: 2112.00054 [cs.CV]. URL: https://arxiv.org/abs/2112.00054.

[8] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. *Adding Conditional Control to Text-to-Image Diffusion Models*. 2023. arXiv: 2302.05543 [cs.CV]. URL: https://arxiv.org/abs/2302.05543.

[9] Nataniel Ruiz et al. *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*. 2023. arXiv: 2208.12242 [cs.CV]. URL: https://arxiv.org/abs/2208.12242.

[10] Scott Geng, Ranjay Krishna, and Pang Wei Koh. "Training with Real instead of Synthetic Generated Images Still Performs Better". In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024*. 2024. URL: https://openreview.net/forum?id=7YnS0mYl8s.

[11] Louis Hémadou et al. "Beyond Internet Images: Evaluating Vision-Language Models for Domain Generalization on Synthetic-to-Real Industrial Datasets". In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024*. 2024. URL: https://openreview.net/forum?id=BgpApqspGw.

[12] Yixuan Li et al. "MatrixCity: A Large-scale City Dataset for City-scale Neural Rendering and Beyond". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2023, pp. 3205–3215.

[13] Ashish Shrivastava et al. *Learning from Simulated and Unsupervised Images through Adversarial Training*. 2017. arXiv: 1612.07828 [cs.CV]. URL: https://arxiv.org/abs/1612.07828.

[14] Robin Rombach et al. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2022. arXiv: 2112.10752 [cs.CV]. URL: https://arxiv.org/abs/2112.10752.

[15] Vajira Thambawita et al. *SinGAN-Seg: Synthetic training data generation for medical image segmentation*. Ed. by Ruxandra Stoean. DOI: 10.1371/journal.pone.0267976. URL: http://dx.doi.org/10.1371/journal.pone.0267976.

[16] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. 2018. arXiv: 1611.07004 [cs.CV]. URL: https://arxiv.org/abs/1611.07004.

[17] Han Zhang et al. *StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks*. 2017. arXiv: 1612.03242 [cs.CV]. URL: https://arxiv.org/abs/1612.03242.

[18] Martin Arjovsky, Soumith Chintala, and Léon Bottou. *Wasserstein GAN*. 2017. arXiv: 1701.07875 [stat.ML]. URL: https://arxiv.org/abs/1701.07875.

[19] Prafulla Dhariwal and Alex Nichol. *Diffusion Models Beat GANs on Image Synthesis*. 2021. arXiv: 2105.05233 [cs.LG]. URL: https://arxiv.org/abs/2105.05233.

[20] Christoph Schuhmann et al. *LAION-5B: An open large-scale dataset for training next generation image-text models*. 2022. arXiv: 2210.08402 [cs.CV]. URL: https://arxiv.org/abs/2210.08402.

[21] Alec Radford et al. *Learning Transferable Visual Models From Natural Language Supervision*. 2021. arXiv: 2103.00020 [cs.CV]. URL: https://arxiv.org/abs/2103.00020.

[22] Chong Mou et al. "T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models". In: *arXiv preprint arXiv:2302.08453* (2023).

[23] Tim Brooks et al. "Video generation models as world simulators". In: *openai.com* (2024). URL: https://openai.com/research/video-generation-models-as-world-simulators.

[24] Jay Zhangjie Wu et al. "Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2023, pp. 7623–7633.

[25] Jamie Shotton et al. "Efficient Human Pose Estimation from Single Depth Images". In: *IEEE transactions on pattern analysis and machine intelligence* 35 (Dec. 2013), pp. 2821–40. DOI: 10.1109/TPAMI.2012.241.

[26] Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman. "Synthetic Data for Text Localisation in Natural Images". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[27] Adrien Gaidon et al. *Virtual Worlds as Proxy for Multi-Object Tracking Analysis*. 2016. arXiv: 1605.06457 [cs.CV]. URL: https://arxiv.org/abs/1605.06457.

[28] Zhenyu Wu et al. "Synthetic Data Supervised Salient Object Detection". In: *Proceedings of the 30th ACM International Conference on Multimedia*. MM '22. ACM, Oct. 2022. DOI: 10.1145/3503161.3547930. URL: http://dx.doi.org/10.1145/3503161.3547930.

[29] Lijun Wang et al. "Learning to Detect Salient Objects with Image-level Supervision". In: *CVPR*. 2017.

[30] Weijia Wu et al. *DiffuMask: Synthesizing Images with Pixel-level Annotations for Semantic Segmentation Using Diffusion Models*. 2024. arXiv: 2303.11681 [cs.CV]. URL: https://arxiv.org/abs/2303.11681.

[31] Weijia Wu et al. *DatasetDM: Synthesizing Data with Perception Annotations Using Diffusion Models*. 2023. arXiv: 2308.06160 [cs.CV]. URL: https://arxiv.org/abs/2308.06160.

[32] Manlin Zhang et al. *DiffusionEngine: Diffusion Model is Scalable Data Engine for Object Detection*. 2023. arXiv: 2309.03893 [cs.CV]. URL: https://arxiv.org/abs/2309.03893.

[33] Yonglong Tian et al. *Learning Vision from Models Rivals Learning Vision from Data*. 2023. arXiv: 2312.17742 [cs.CV]. URL: https://arxiv.org/abs/2312.17742.

[34] Amirhossein Kazerouni et al. *Diffusion Models for Medical Image Analysis: A Comprehensive Survey*. 2023. arXiv: 2211.07804 [eess.IV]. URL: https://arxiv.org/abs/2211.07804.

[35] Man M. Ho et al. "DISC: Latent Diffusion Models with Self-Distillation from Separated Conditions for Prostate Cancer Grading". In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024*. 2024. URL: https://openreview.net/forum?id=NcCAy6PXNF.

[36] Doan Thinh Vo. "An Approach to Synthesize Thermal Infrared Ship Images". In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024*. 2024. URL: https://openreview.net/forum?id=XQQ12dIR7C.

[37] Yasser Benigmim et al. "One-Shot Unsupervised Domain Adaptation With Personalized Diffusion Models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2023, pp. 698–708.

[38] Donggeun Ko et al. "DiffInject: Revisiting Debias via Synthetic Data Generation using Diffusion-based Style Injection". In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024*. 2024. URL: https://openreview.net/forum?id=jSB5wlUU3p.

[39] Laith Alzubaidi, Jian Bai, Ali Al-Sabaawi, et al. "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications". In: *Journal of Big Data* 10.46 (2023). DOI: 10.1186/s40537-023-00727-2. URL: https://doi.org/10.1186/s40537-023-00727-2.

[40] Jiaxi Wu, Jiaxin Chen, and Di Huang. *Entropy-based Active Learning for Object Detection with Progressive Diversity Constraint*. 2022. arXiv: 2204.07965 [cs.CV]. URL: https://arxiv.org/abs/2204.07965.

[41] Maximilian Dreyer et al. *Revealing Hidden Context Bias in Segmentation and Object Detection through Concept-specific Explanations*. 2022. arXiv: 2211.11426 [cs.CV]. URL: https://arxiv.org/abs/2211.11426.

[42] Niklas Bunzel, Nicolas Göller, and Raphael Antonius Frick. "Identifying and Generating Edge Cases". In: *Proceedings of the 2nd ACM Workshop on Secure and Trustworthy Deep Learning Systems*. SecTL '24. Singapore, Singapore: Association for Computing Machinery, 2024, 16–23. ISBN: 9798400706912. DOI: 10.1145/3665451.3665529. URL: https://doi.org/10.1145/3665451.3665529.

[43] Tsung-Yi Lin et al. "Focal Loss for Dense Object Detection". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2999–3007. DOI: 10.1109/ICCV.2017.324.

[44] Nicholas Konz et al. *Anatomically-Controllable Medical Image Generation with Segmentation-Guided Diffusion Models*. 2024. arXiv: 2402.05210 [eess.IV]. URL: https://arxiv.org/abs/2402.05210.

[45] Setareh Dabiri et al. *Realistically distributing object placements in synthetic training data improves the performance of vision-based object detection models*. 2023. arXiv: 2305.14621 [cs.CV]. URL: https://arxiv.org/abs/2305.14621.

[46] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*. 2015. arXiv: 1405.0312 [cs.CV]. URL: https://arxiv.org/abs/1405.0312.

[47] Marius Cordts et al. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. 2016. arXiv: 1604.01685 [cs.CV]. URL: https://arxiv.org/abs/1604.01685.

[48] Gui-Song Xia et al. *DOTA: A Large-scale Dataset for Object Detection in Aerial Images*. 2019. arXiv: 1711.10398 [cs.CV]. URL: https://arxiv.org/abs/1711.10398.

[49] Holger Caesar et al. *nuScenes: A multimodal dataset for autonomous driving*. 2020. arXiv: 1903.11027 [cs.LG]. URL: https://arxiv.org/abs/1903.11027.

[50] Christian Cianfarani et al. *Understanding Robust Learning through the Lens of Representation Similarities*. 2022. arXiv: 2206.09868 [cs.LG]. URL: https://arxiv.org/abs/2206.09868.

[51] Xuezhou Zhang, Xiaojin Zhu, and Stephen J. Wright. *Training Set Debugging Using Trusted Items*. 2018. arXiv: 1801.08019 [cs.LG]. URL: https://arxiv.org/abs/1801.08019.

[52] Chenhongyi Yang, Lichao Huang, and Elliot J. Crowley. *Plug and Play Active Learning for Object Detection*. 2024. arXiv: 2211.11612 [cs.CV]. URL: https://arxiv.org/abs/2211.11612.

[53] Haichao Yu et al. *High-Resolution Deep Image Matting*. 2021. arXiv: 2009.06613 [cs.CV]. URL: https://arxiv.org/abs/2009.06613.

[54] Tuan Pham Minh et al. *Active Label Refinement for Semantic Segmentation of Satellite Images*. 2023. arXiv: 2309.06159 [cs.CV]. URL: https://arxiv.org/abs/2309.06159.

[55] Yueyuan Li et al. "Choose Your Simulator Wisely: A Review on Open-Source Simulators for Autonomous Driving". In: *IEEE Transactions on Intelligent Vehicles* 9.5 (May 2024), 4861–4876. ISSN: 2379-8858. DOI: 10.1109/tiv.2024.3374044. URL: http://dx.doi.org/10.1109/TIV.2024.3374044.

[56] Jinlong Li et al. *Domain Adaptation based Object Detection for Autonomous Driving in Foggy and Rainy Weather*. 2024. arXiv: 2307.09676 [cs.CV]. URL: https://arxiv.org/abs/2307.09676.

[57] Shuai Shao et al. "CrowdHuman: A Benchmark for Detecting Human in a Crowd". In: *arXiv preprint arXiv:1805.00123* (2018).

[58] Haroon Idrees et al. *Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds*. 2018. arXiv: 1808.01050 [cs.CV]. URL: https://arxiv.org/abs/1808.01050.