Evaluation of 3D-LiDAR based person detection algorithms for edge computing

Dennis Basile[1], Dennis Sprute[1], Helene Dörksen[2], and Holger Flatt[1]

[1] Fraunhofer IOSB, Industrial Automation Branch (IOSB-INA), Campusallee 1, 32657 Lemgo
[2] OWL University of Applied Sciences and Arts, Campusallee 2, 32657 Lemgo

Abstract This paper addresses the need for reliable person detection systems in public spaces by developing a novel dataset tailored for solid-state 3D-LiDAR sensors and evaluating various neural network architectures. The dataset was created using a Blickfeld solid-state 3D-LiDAR, capturing 265 point clouds in a controlled test environment modeled on a three-lane pedestrian crossing. The neural network architectures evaluated include VoxelNeXt, PillarNet, SECOND, PointPillar, CenterPoint, Voxel-R-CNN, PointRCNN, PartA2, and PV-RCNN. The evaluation methodology follows the KITTI benchmark metric for performance analysis. Key results indicate that voxel-based approaches like SECOND and VoxelNeXt achieve inference speeds of 10.3 FPS and 9.8 FPS on an NVIDIA Jetson AGX platform, respectively, with mean Average Precision (mAP) scores of 95% and 90%. In contrast, the hybrid approach PV-RCNN, which combines voxel-based and point-based methods, achieves a mAP of 92% but a slower inference speed of 2.5 FPS. These results underscore the trade-offs between speed and accuracy in person detection using solid-state 3D-LiDAR, highlighting the potential of voxel-based methods for real-time applications. The results contribute to the advancement of person detection technologies in public security and smart city initiatives.

Keywords 3D-LiDAR, person detection, edge computing

DOI: 10.58895/ksp/1000174496-14 erschienen in:
Forum Bildverarbeitung 2024
DOI: 10.58895/ksp/1000174496 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000174496/

1 Introduction

The increasing demand for robust and reliable person detection systems in public spaces has driven advancements in sensor technology and machine learning algorithms. Accurate detection is crucial for applications like public security, traffic management, and smart city initiatives. In the domain of public space surveillance, these systems must accurately localize and classify objects in real-time and operate under challenging conditions such as fog, snow, and rain, while complying with the General Data Protection Regulation (GDPR) in the European Union. Existing systems use various sensors like PIR, laser barriers, radar, and cameras. However, each technology has drawbacks. For example, PIR sensors struggle with detecting groups due to lack of a classical field of view, while cameras, although effective with AI for detection and classification, raise privacy concerns under EU-GDPR [1].

In contrast, solid-state 3D-LiDAR technology shows great potential by generating precise 3D point clouds for privacy-preserving and reliable detection [2]. This makes 3D-LiDAR ideal for applications requiring accuracy, real-time operation, environmental resilience, and data privacy. Currently, 3D-LiDAR is extensively used and researched in autonomous driving systems [3]. However, the available datasets for training neural networks focus on automotive use and may not encompass the broader range of potential applications. They are captured with rotating 3D-LiDAR sensors, whose characteristics, such as resolution, range, and field of view, differ significantly from solid-state 3D-LiDARs. Transferring an existing dataset to the characteristics of a solid-state 3D-LiDAR is challenging. Consequently, there is no sufficient dataset for independent analysis using solid-state 3D-LiDAR sensors. This necessitates the creation of new datasets targeting the specific hardware characteristics of solid-state 3D-LiDAR to achieve optimal performance in people detection with deep learning approaches. Furthermore, there has been no comprehensive comparison of neural network architectures with respect to the specific requirements for person detection in public spaces using solid-state 3D-LiDAR sensors and edge computing. Therefore, this paper contributes by developing a novel dataset for solid-state 3D-LiDAR sensors and performing a thorough comparison of various neural network architectures addressing the requirements for person detection systems in public environments.

The rest of this paper is organized as follows: Section II reviews related work focusing on 3D-LiDAR datasets. Section III describes the generation of the novel dataset based on a design flow and a person classification scheme. In Section IV, the approach is applied within a case study by creating a dataset used to train different CNN architectures. Section V presents the results of the evaluation, and finally, Section VI concludes the paper.

2 Related Work

3D-LiDAR technology has become crucial for advanced driver assistance systems, primarily used for detecting obstacles [4]. Current implementations mainly utilize rotating 3D-LiDARs, as demonstrated by datasets like KITTI, which is a standard benchmark in this field [5]. Several other datasets have been created (refer to Tab. 1), all based on rotating 3D-LiDARs. These datasets are primarily designed for automotive applications, potentially limiting their broader applicability.

Table 1: Overview of various LiDAR datasets.

Dataset	LiDAR Type	LiDAR System	Licensing
KITTI [6]	Rotating	Velodyne HDL-64E	Non-commercial
Waymo Open Dataset [7]	Rotating	In-house development	Non-commercial
nuScenes [8]	Rotating	Velodyne HDL-32E	Non-commercial
PandaSet [9]	Rotating	Hesai Pandar64	Commercial
Argoverse 2 [10]	Rotating	Velodyne VLP-32C	Non-commercial
ONCE [11]	Rotating	40-Beam LiDAR	Non-commercial

Solid-state LiDARs, however, offer several advantages over rotating LiDARs, such as being more compact, lighter, more energy- and cost-efficient. Additionally, without mechanical components, they are maintenance-free and have a longer lifespan [12]. Recent studies suggest that solid-state LiDARs can also be used effectively for tasks beyond obstacle detection, like pedestrian recognition. For example, Peng et al. [13] explored using solid-state LiDAR and cameras for pedestrian detection. However, using camera data raises privacy concerns as it can capture identifiable personal information.

Sprute et al. [14] address the challenge of achieving high-resolution spatial coverage with solid-state LiDAR without cameras, focusing on detecting people using deep learning techniques. 3D-LiDAR sensors capture point clouds, which are then converted into depth images through clustering techniques. Afterwards, they are processed with a ResNet-based neural network for object classification. This method is computationally intensive, limiting real-time processing on embedded systems. While it improves spatial coverage and detection accuracy, it does not offer direct real-time processing of point clouds, which can be a limitation in scenarios requiring immediate feedback.

Several points from current research highlight the need for further investigation. First, detecting people using solid-state LiDAR and deep learning is feasible, but existing datasets are designed for rotating LiDAR systems, limiting their applicability. A new dataset for solid-state LiDAR is needed.

Second, direct processing of point clouds for person detection is rarely explored. Most studies convert point clouds into depth images before classification, which is computationally demanding and unsuitable for real-time applications.

Third, embedded systems have not been sufficiently considered. Mapping deep learning architectures onto embedded systems could enhance efficiency and applicability, especially for compact, energy-efficient use cases.

This work develops a new dataset for solid-state LiDAR and evaluates deep learning architectures for direct point cloud processing. The aim is to identify effective deep learning models for implementation on embedded systems for efficient person detection.

3 Novel dataset for person detection

Since there are currently no publicly available datasets specifically tailored to the requirements for person detection using solid-state LiDAR, a custom dataset for model training is required.

A solid-state 3D-LiDAR sensor system is employed to capture data, mapping the surroundings as a 3D point cloud. The orientation and position of the sensor remain static throughout the data collection process, ensuring consistent raw data acquisition.

To create a dataset, several processing steps must be carried out. The raw data has to be stored, followed by storing the raw data in individual frames. These frames then have to be normalized and converted into a point cloud format. Subsequently, the LiDAR coordinate data has to be adjusted to meet the specific requirements for training. After this, the data has to be annotated, and labels have to be created. Finally, the dataset has to be split into training, validation and test subsets.

The sensor setup was established in a specially designed test environment on the premises of the Fraunhofer IOSB-INA Institute, ensuring unobstructed visibility. The setup is based on previous work of Sprute et al. [15]. The LiDAR sensor was installed at a height of four meters with a 16° tilt to ensure optimal coverage of the entire area. The setup was focused on a distance of 9 meters and was directed towards a three-lane pedestrian crossing at an intersection. The data was captured using a solid-state 3D LiDAR sensor from the company Blickfeld [16]. The sensor was configured with a field of view of 72° x 30°, a framerate of 2.4 Hz, and 200 scan lines.

The dataset is collected from the recorded raw data, where different individuals passed by the LiDAR within a range of up to 30 meters. For the training of the deep learning algorithms, a single class 'Person' with different variations was considered. This ensured that the model could detect and analyze various person types and their movement patterns. The manual annotation of the single objects was carried out carefully, as it directly impacts the quality of the detection results after training. The entire dataset consists of 265 different point clouds. An example of the classes annotated in the dataset can be seen in Fig. 1.

To ensure the versatility and robustness of the proposed recognition system, various classes of people based on their relevance in public space have to be provided [15]. The following classes are used to extend the dataset: 1) individuals without physical disabilities, 2) individuals with forearm crutches, 3) individuals with rollators, 4) individuals with mobile phones, 5) groups of people, and 6) individuals with walking sticks.

These annotations reflect common situations in public areas, capturing a wide range of human activities and interactions. Recognizing such diverse scenarios is particularly relevant for surveillance and public safety applications, enhancing the model's ability to detect individuals accurately in various contexts. This variety of annotated classes ensures that the developed model is capable of recognizing and correctly classifying different situations and groups of people.

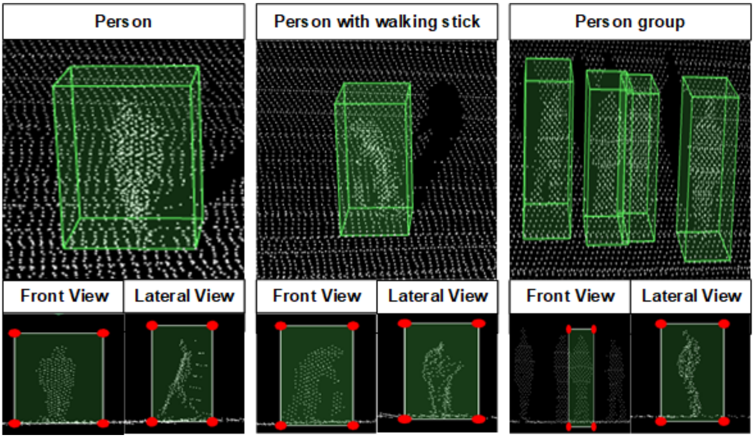

Figure 1: Examples from the custom dataset of manually annotated people.

4 Case Study

4.1 Neural Network

There are various approaches of deep learning architectures for the direct processing of point clouds, which can be categorized into voxel-based, point-based, and hybrid methods.

Voxel-Based Approaches

Voxel-based approaches partition the point cloud into small 3D cubes (voxels) and extract features from each voxel using a Voxel Feature Encoding (VFE) layer. These methods convert the irregular point cloud data into a regular grid, which can then be processed using sparse convolutional neural networks (CNNs) [17]. Advantages of these methods include fast inference times and reduced computational load. However, there are drawbacks, such as information loss due to the choice of voxel size [18]. Examples of voxel-based approaches used in this study include: (1) SECOND [17], (2) PointPillars [19], (3) PillarNet [20], (4) CenterPoint [21], (5) VoxelNeXt [22], (6) PartA2 [23], (7) Voxel-R-CNN [24].

Point-Based Approaches

Point-Based approaches directly process the point cloud . These method use PointNet++ [25], to learn features directly from the raw points, achieving a higher level of detail. However, they often incur higher computational costs due to the unstructured nature of the data, increased memory usage, and slower inference speeds [26]. Example of a point-based approach used in this study is PointRCNN [27]

Hybrid-Based Approaches

The hybrid method is an extension that combines voxel- and point-based approaches to point cloud processing, combining the strengths of each. Voxel-based methods are faster but can lose information, while point-based methods retain all information but are slower to process. This hybrid approach attempts to combine efficient computation with comprehensive data representation. An example of a hybrid-based approach used in this study is PV-RCNN [28]

4.2 Training

The open-source framework Point Cloud Detection (OpenPCDet) [29] was employed for training and execution of the deep learning architectures described in Section 4.1. The deep learning architectures were trained on a Windows system with the following specifications: 64 GB of DDR4 RAM, an AMD Ryzen 9-3900X 12-core processor, and an RTX2080 graphics card with 8 GB of memory. To ensure the reliability of the results, the dataset was randomly divided into two distinct sets: 70% for training and 30% for validation. To enhance the performance of the trained model, data augmentation techniques were employed to artificially expand the dataset [30]. These techniques included rotation, scaling, and mirroring of the point cloud, as well as the generation of additional bounding boxes and their point data based on the training dataset through the introduction of artificial elements.

The extended Adam algorithm, OneCycleLR [31], was employed for all architectures for the optimization of the neural network's weights mentioned in Section 4.1, wherein a variable learning rate was utilized during training. A maximum learning rate of 10^{-4} was selected, with a momentum of 0.95–0.85. The training process was performed with batch sizes of 6 point clouds over 120 epochs.

5 Results

The calculation to analyze performance is based on the KITTI benchmark procedure [5]. Nine distinct neural network architectures were trained on our novel dataset and subsequently evaluated in terms of their performance, including measures such as average precision (AP) and inference time. In Tab. 2, the results of the conducted investigation of the evaluated deep learning architectures with the custom dataset for AP and the measured inference time on an edge computing device Nvidia Jetson AGX system are presented.

The results demonstrate that voxel-based approaches, such as SECOND, VoxelNeXt, or Voxel-R-CNN, achieve notable performance in both AP and inference time, offering a suitable balance between speed and accuracy when compared to point-based and hybrid approaches. These results are significantly better when compared to the performance of a point-based approach, such as PointRCNN and a hybrid approach, such as PV-RCNN. Some qualitative detection results are shown in Fig. 2.

Table 2: Comparison of architecture performance.

Type	Architecture	AP (IoU = 0.5)	Inference time (FPS)
Voxel	CenterPoint	0.92	7.7
Voxel	Part-A2	0.95	5.0
Voxel	PillarNet	0.91	7.2
Voxel	PointPillar	0.89	9.1
Voxel	SECOND	0.95	**10.3**
Voxel	Voxel-R-CNN	**0.97**	7.2
Voxel	VoxelNeXt	0.90	9.8
Point	PointRCNN	0.92	1.6
Voxel/Point	PV-RCNN	0.90	2.5

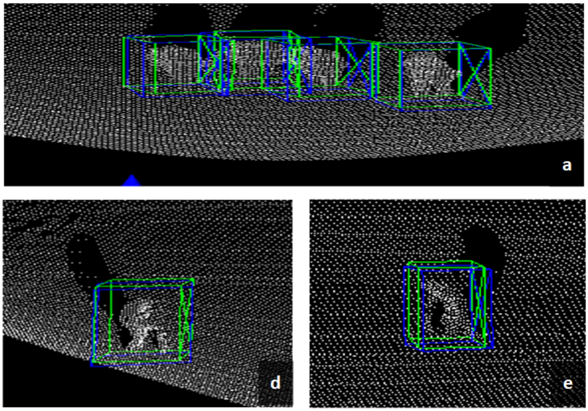

Figure 2: Exemplary person detection based on SECOND architecture. The blue rectangles represent the reference bounding boxes, while the green rectangles indicate the predicted bounding boxes from the neural network.

6 Conclusions and Future Work

The study employing the newly created dataset demonstrates that voxel-based methods, particularly SECOND, achieved the best results, reaching 10.3 FPS with an average precision (AP) of 95%. This indicates that classification and localization using point clouds collected with solid-state 3D-LiDAR sensor are possible with an embedded system like the Nvidia Jetson AGX. The evaluation of nine deep learning algorithms for processing 3D point clouds with a solid-state 3D-LiDAR sensor on an edge computing system revealed that single-stage methods based on voxel preprocessing are most effective. Specifically, SECOND, VoxelNeXt, and PointPillar showed high classification and localization performance with real-time processing capabilities. These results confirm that appropriate voxel-based deep learning architectures exist to implement a person detection system on an edge computing platform with a solid-state 3D-LiDAR sensor, enabling efficient real-time person detection and visualization of 3D point clouds.

Future work will focus on refining the dataset to include more diverse and realistic point cloud scenes, addressing variations in weather conditions and background objects. Class separation and the inclusion of new classes, such as people with bicycles and strollers, will be investigated to enhance the system's robustness and flexibility. Finally, the approach will be integrated into an embedded smart sensor system, designed for usage in public spaces.

References

1. A.-M. C. Drăgulinescu, I. Marcu, S. Halunga, and O. Fratu, "Persons Counting and Monitoring System Based on Passive Infrared Sensors and Ultrasonic Sensors (PIRUS)," in *Pervasive Computing Paradigms for Mental Health*. Springer International Publishing, 2018, vol. 207, pp. 100–106.
2. N. Li, C. P. Ho, J. Xue, L. W. Lim, G. Chen, Y. H. Fu, and L. Y. T. Lee, "A progress review on solid-state lidar and nanophotonics-based lidar sensors," *Laser & Photonics Reviews*, vol. 16, no. 11, p. 2100511, 2022.
3. K. Li and L. Cao, "A Review of Object Detection Techniques," in *International Conference on Electromechanical Control Technology and Transportation (ICECTT)*. IEEE, May 2020, pp. 385–390.
4. Y. Li and J. Ibanez-Guzman, "Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, Jul. 2020.
5. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3354–3361.
6. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.
7. P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2443–2451.
8. H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, jun 2020, pp. 11 618–11 628.
9. P. Xiao, Z. Shao, S. Hao, Z. Zhang, X. Chai, J. Jiao, Z. Li, J. Wu, K. Sun, K. Jiang, Y. Wang, and D. Yang, "PandaSet: Advanced Sensor Suite Dataset

for Autonomous Driving," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Sep. 2021, pp. 3095–3101.

10. B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, D. Ramanan, P. Carr, and J. Hays, "Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting," *arXiv*, 2023, version Number: 1.
11. J. Mao, M. Niu, C. Jiang, H. Liang, J. Chen, X. Liang, Y. Li, C. Ye, W. Zhang, Z. Li, J. Yu, H. Xu, and C. Xu, "One Million Scenes for Autonomous Driving: ONCE Dataset," *arXiv*, 2021, version Number: 3.
12. Y. Li and H. Shi, Eds., *Advanced driver assistance systems and autonomous vehicles: from fundamentals to applications*. Springer, 2022.
13. Z. Peng, Z. Xiong, Y. Zhao, and L. Zhang, "3-d objects detection and tracking using solid-state lidar and rgb camera," *IEEE Sensors Journal*, vol. 23, no. 13, pp. 14 795–14 808, 2023.
14. D. Sprute, T. Westerhold, F. Hufen, H. Flatt, and F. Gellert, "DSGVO-konforme Personendetektion in 3D-LiDAR-Daten mittels Deep Learning Verfahren (in German)," in *Bildverarbeitung in der Automation: Ausgewählte Beiträge des Jahreskolloquiums BVAu 2022*. Springer, 2023, pp. 33–45.
15. D. Sprute, F. Hufen, T. Westerhold, and H. Flatt, "3D-LiDAR-based pedestrian detection for demand-oriented traffic light control," in *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*, 2023, pp. 1–7.
16. Blickfeld GmbH, "Technical manual of the blickfeld lidar cube 1," *https://www.blickfeld.com*, 2021. [Online]. Available: https://www.blickfeld.com/wp-content/uploads/2022/10/Blickfeld-A5-Manual_en_v.4.2.pdf,[Onlineaccess:2024-02-23]
17. Y. Yan, Y. Mao, and B. Li, "SECOND: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, 2018.
18. M. Ye, S. Xu, and T. Cao, "Hvnet: Hybrid voxel network for lidar based 3d object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1628–1637.
19. A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 689–12 697.
20. G. Shi, R. Li, and C. Ma, "Pillarnet: Real-time and high-performance pillar-based 3d object detection," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Springer Nature Switzerland, 2022, pp. 35–52.

21. T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3D object detection and tracking," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2021, pp. 11 779–11 788.
22. Y. Chen, J. Liu, X. Zhang, X. Qi, and J. Jia, "VoxelNeXt: Fully sparse voxelnet for 3D object detection and tracking," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 21 674–21 683.
23. S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li, "From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 08, pp. 2647–2664, 2021.
24. J. Deng, S. Shi, P. Li, W. Zhou, Y. Zhang, and H. Li, "Voxel R-CNN: Towards high performance voxel-based 3D object detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 1201–1209, 05 2021.
25. C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.
26. M. Drobnitzky, J. Friederich, B. Egger, and P. Zschech, "Survey and systematization of 3D object detection models and methods," *The Visual Computer*, vol. 40, no. 3, pp. 1867–1913, 2024.
27. S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2019, pp. 770–779.
28. S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "Pv-rcnn: Point-voxel feature set abstraction for 3d object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10 526–10 535.
29. O. D. Team, "Openpcdet: An open-source toolbox for 3d object detection from point clouds," https://github.com/open-mmlab/OpenPCDet, 2020.
30. C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019.
31. L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," SPIE, pp. 369–386, 2019.