Al-Quraan, M., Khan, A. R., Mohjazi, L. , Centeno, A. , Zoha, A. and Imran, M. A. (2022) A Hybrid Data Manipulation Approach for Energy and Latency-Efficient Vision-Aided UDNs. In: 2021 Eighth International Conference on Software Defined Systems (SDS), Gandia, Spain (Virtual), 6-9 Dec 2021, ISBN 9781665458207

(doi: 10.1109/SDS54264.2021.9732115)

A Hybrid Data Manipulation Approach for Energy and Latency-Efficient Vision-Aided UDNs

Mohammad Al-Quraan, Ahsan Raza Khan, Lina Mohjazi, Anthony Centeno, Ahmed Zoha, and Muhammad Ali Imran

James Watt School of Engineering, University of Glasgow, Glasgow, UK

Email: {m.alquraan.1, a.khan.9}@research.gla.ac.uk,

{Lina.Mohjazi, Anthony.Centeno, Ahmed.Zoha, Muhammad.Imran}@glasgow.ac.uk

***Abstract*—The combination of deep learning (DL) and computer vision (CV) is shaping the future of wireless communications by supporting the operations of ultra-dense networks (UDNs). However, vision-aided wireless communications (VAWC) are highly dependent on DL algorithms that rely on a wide range of multimodal data stored at a central location. Although the performance of the DL model is improved when the model becomes deeper, the need for a large number of datasets for model training incurs more computational complexity in terms of model training time and storage size. Hence, the energy efficiency of the network will become worse due to the higher energy costs associated with model training and transmitting a large amount of data over wireless links. Therefore, a critical challenge is to reduce the computational complexity and bandwidth utilisation of DL-based vision-aided UDNs without compromising their performance. In this paper, we adopt single-channel (SICH) images, joint photographic expert group (JPEG) image compression (COMP), and object detection (ODET) to form a hybrid data manipulation technique. This technique can reduce the model computation cost and data storage volume, as well as alleviate the transmission burden on the wireless links to make future wireless networks more reliable and energy efficient. Specifically, these techniques are used to manipulate datasets before using them in model training. Compared to reference datasets, simulation results show that our hybrid technique that combines SICH, COMP, and ODET achieves the best performance in reducing the model computation complexity and network transmission latency and improving network's energy efficiency. This results in a computation time improvement of 34%, a significant reduction of 86% in memory size for data storage, reducing data transmission time by 83%, and 82.5% more energy efficient networks.**

***Index Terms*—Vision processing, energy efficiency, transmission latency, deep learning, computer vision, wireless communications, data compression, 5G and beyond.**

I. INTRODUCTION

Fifth-generation (5G) and beyond-5G (B5G) wireless networks are characterised by the adoption of high operating frequencies, i.e., utilising millimetre wave (mmWave) and terahertz (THz) technologies [1]. The reliance on higher frequency bands brought advantages in terms of high data rates, massive device connectivity, and ultra-reliable and low latency communication. Moreover, the mmWave and THz bands coupled with beamforming and massive multiple-input multiple-output (MIMO) technologies shift the paradigm of wireless networks from large omnidirectional cells to small directional cells, thus, forming the concept of ultra-dense networks (UDNs). These networks can serve a tremendous number of connected devices and realise many revolutionary applications, such as autonomous driving, augmented reality, and video surveillance in smart cities. Consequently, enormous amounts of data will be generated, both at the core and the edge of the network [2].

Deep learning (DL) has shown its importance in handling large volumes of datasets to identify hidden trends and patterns [3]. Hence, the use of DL in UDNs has attracted many research interests to improve their operation, including resource allocation [4], channel estimation and tracking [5], beam selection [6], among others. However, due to the dependence on higher frequency bands, UDNs encounter some critical challenges that cannot be addressed using traditional network operation, even when employing DL algorithms. For instance, mmWave and THz signals suffer from severe attenuation and penetration losses, and can be easily blocked by objects located between the transmission points. Beam blockage is a non-trivial challenge that the UDNs are facing, especially in dynamic environments. Therefore, an emerging approach of combining DL with computer vision (CV) has recently emerged [7].

Vision-aided wireless communications (VAWC) use visual sensory and wireless channel data to address complex problems in UDNs [8]. The fusion of DL and CV is expected to tackle critical challenges, such as beam selection, blockage prediction, and handover. Moreover, VAWC exploits visual sensory information like red, green, blue (RGB)/depth images, videos, light detection and ranging (LiDAR), and 3D point cloud collected from the target area by cameras/scanners. The new mobile edge computing (MEC) paradigm introduced by the European telecommunications standards institute (ETSI) brings cloud computing capabilities to the edge of the radio access network. Following the conventional approach of training machine learning (ML) models in a central location, i.e., an edge server, VAWC datasets are sent to the central server through wireless links to train DL models designed to tackle challenging problems in UDNs. However, DL models require large quantities of datasets and training rounds to perform efficiently. Consequently, transmitting a large amount of data over the wireless links requires a significant amount of energy [9], resulting in a substantial network carbon footprint and a considerable strain on the wireless communication links, besides incurring high transmission delay and the need for large storage size. In addition, the underlying mathematical operations of DL algorithms yield heavy computational burdens, especially for deeper DL models, which consume significant amounts of energy. These challenges are exacerbated when handling multimodal information of VAWC systems.

From an environmental perspective and given that most global energy usage nowadays hugely depends on fossil fuels, the increase in energy consumption means increased levels of carbon dioxide equivalent (CO_2e) emissions, which constitutes the main reason for global warming and climate change. Human activities caused 1.0°C global warming by 2020, and it is estimated to reach 1.5°C between 2032 and 2050 if the increase rate remains the same [10]. Many global endeavors have been undertaken to prevent such a rise, which can reach the point of no return and cause catastrophic natural disasters. Therefore, it is paramount to reconsider the energy consumption patterns.

Capitalising on the above, the main focus of this paper is to improve DL-based vision-aided systems in UDNs to be more energy efficient by reducing the computational complexity of DL models represented by the time of model training and data storage. Additionally, the energy efficiency is further improved by decreasing the energy cost of data transmission over the wireless links, along with significantly decreasing the network transmission latency. After surveying the literature, we noticed that almost all the presented works that are concerned with energy efficiency problems in DL algorithms focus on modifying the model settings through its structure [11], [12] or parameters [13], [14]. Although these works have shown an efficient model computation, they also cause a degradation in the prediction accuracy.

To the best of our knowledge, no prior works have studied the energy efficiency of VAWC systems. Also, no studies alleviate the computation complexity of DL models through modulating the input datasets in lieu of modifying the model setting. In this paper, we propose a hybrid data manipulation approach by combining object detection (ODET), single-channel (SICH) images, and joint photographic experts group (JPEG) image compression (COMP) techniques to reduce the computational complexity of DL-based vision aided systems by modulating the visual sensory training dataset. Specifically, we reduce the memory needed to store the massive amount of visual information by utilising COMP and ODET techniques. Furthermore, we also reduce the model training and inference time by eliminating redundant information in input data using SICH images. The proposed techniques can significantly enhance energy efficiency by reducing the computational cost of DL models without compromising the system's overall performance. Data transmission time over wireless links is also improved, which is essential for realising time-critical applications. The following points demonstrate our main contributions:

- We present a hybrid data manipulation approach that minimises the computational complexity of DL models in vision-aided systems by modulating the input datasets instead of changing the model structure.
- We improve the energy efficiency of DL-based vision aided systems in UDNs by reducing the network's carbon footprint associated with the high energy cost needed to perform model training and data transmission.
- The work also alleviates the burden of data transmission on wireless links and significantly minimises data transmission latency, which is critical for time-sensitive applications.

The rest of this paper is organised as follows. Section II discusses the integration of DL with CV and the impact of modern DL models on the computation cost. In Section III, we introduce the proposed techniques that reduce the computation complexity of DL-based CV models. Simulation results are shown in Section IV. Finally, Section V gives concluding remarks.

II. DL-Based CV-Aided Systems

DL is the cutting-edge of artificial intelligence (AI). It provides systems with the ability to automatically learn how to give accurate predictions without human intervention. Unlike basic ML techniques that rely on complex feature engineering and engineer guidance, DL models can make intelligent decisions without the need for data analysis. Deeper DL models that contain multiple hidden layers have proven to be important through achieving breakthroughs in a wide variety of fields such as gaming, speech, and vision [15]. Moreover, the combination of DL and CV has the potential to improve the operation of wireless communication systems. This fusion is realised due to the fact that the mmWave and THz frequency bands will make UDNs highly dependent on line-of-sight (LoS) communications, which are aligned with the field-of-view of vision sensors.

A. VAWC Overview

The reliance on the LoS propagation characteristic of high frequency signals is consistent with the visual information captured from cameras/scanners, and the direct view is essential for both. By allowing wireless systems to have a sense of the surrounding environment and exploiting the visual sensory information of the covered area in addition to wireless data, CV is envisioned to play a major role in future wireless networks. Few works have been presented in the literature to demonstrate the efficacy of leveraging CV integrated with DL in wireless communication systems. For instance, the work in [16] proposes a camera-assisted predictive handover algorithm. The depth images taken by the cameras and the measured network throughput are used to train a DL model for link quality prediction. The model learns the relationship between the depth images and measured throughput to estimate future link quality and guide the network to make optimal decisions in advance.

In [17], Klautau et al. utilise the DL algorithm in addition to LiDAR information collected in vehicles for beam prediction. This work aims to reduce the beam selection overhead by relying only on LiDAR information and BS location. The vehicle uses this information to predict the best candidate beams and sends them to the BS to determine the best beam for transmission. In order to predict LoS link blockages in advance, the study in [18] develops a DL model that utilises RGB images and beamforming vectors to predict beam blockage in UDNs proactively. While the work in [19] proposes a deep reinforcement learning (DRL) model that exploits sequences of RGB images to predict beam blockage, hence making efficient handover decisions in advance. According to the aforementioned research studies and their results, it is clear that utilising visual sensory information will be a substantial part of designing and realising robust and reliable wireless communication systems. However, combining DL with CV is accompanied with huge computation and energy consumptions that may hinder their wide deployment in wireless networks. Therefore, this work mainly focuses on proposing data manipulation techniques that can reduce the computational complexity of DL-based CV systems, making them more energy efficient.

B. Energy Concerns of Modern DL Approaches

DL algorithms have undergone a significant shift towards deeper neural networks that require massive quantities of data for training. This is attributed to advances in hardware as well as improved computation techniques. However, the remarkable performance of large neural networks comes at the expense of considerable energy consumption accompanied with the underlying mathematical computations [20]. Therefore, substantial financial and environmental costs are incurred. Modern DL models continue to grow in the number of layers to give better generalisation by learning all intermediate features at different levels of abstraction. For instance, the study in [20] shows that the relationship between DL model size and the amount of computation needed to train these models until convergence has an exponential relationship. This relationship is verified using some of the well-known DL models between 2012 and 2018. Based on this relationship, it is evident that the depth of the new proposed DL models will continue to grow to the point of intolerable computation and energy consumption cost. For this reason, it is imperative to find new approaches that help address the computation complexity issue of DL networks.

III. Proposed DL-Based CV Energy-Efficient Approach

Innovative 5G services and enhanced applications anticipated in sixth-generation (6G) networks render the wireless networks to be more complex than today's paradigm. Hence, the traditional methods used in network design, operation, and optimisation will be inadequate. To make wireless networks intelligent, a data-driven AI-based paradigm should be utilised, so that the network nodes can determine the best policy by exploiting the knowledge extracted from the collected data [21]. However, the newly emerging DL algorithms are scaling rapidly and approaching the computation limit. Continuing this path will soon become unbearable technically, financially, and environmentally. Additionally, the modern approach of

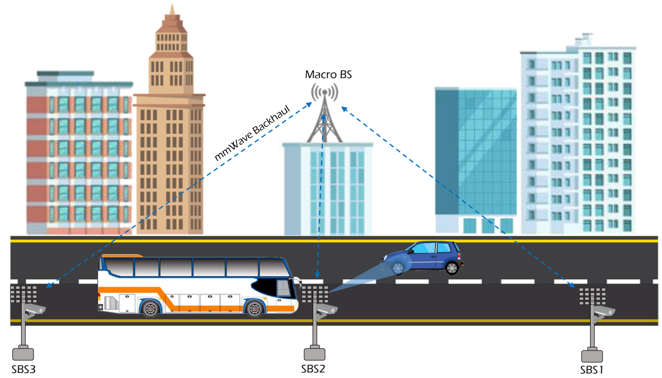


Figure 1: UDN consists of a macro BS and three small BSs each with an attached RGB camera.

exploiting CV in optimising UDNs will further increase the computation complexity and burden the wireless transmission links due to the reliance on big multimodal data. In other words, the ML community will be pushed toward two directions, either by optimising the current DL algorithms and reducing their computation cost or by developing new, less resource-intensive ML technologies. In the following, we will discuss our wireless communication system model and the proposed techniques that will reduce the computation cost of current DL models and make the wireless networks more energy efficient.

A. System Model

In this work, we assume the centralised model training approach in UDNs. Our system considers the UDN that includes one macro BS and three small BSs (SBSs) operating at 60 GHz, as illustrated in Fig. 1. Each SBS contains an array of antennas that enables beamforming technology and serves the users by selecting the optimum beam that provides the highest link budget for single-antenna users using LoS communication. In addition, each SBS has a vision sensor (RGB camera) to obtain a sense of the surrounding environment and assist the operation of the wireless network. The captured vision information is sent to a central server located at the macro BS through mmWave backhaul links. The role of the central server is to receive visual and wireless information from SBSs and employ them in training ML models that can intelligently predict beam blockages, hence improving the performance of wireless communication systems. Moreover, the scenario under study includes a moving user and a stationary blocking object, i.e., bus, that can block the LoS communication between the SBS and the user. For the comparison purpose, we consider the distributed cameras blocked view scenario of the ViWi dataset [22] and adopt the beam blockage problem in [7] as a baseline work. The base work utilises the pretrained ResNet18, a convolutional neural network (CNN) model with 18 hidden layers, where each layer contains a different number of neurones. This model was trained using the ImageNet dataset to classify images based on 1000 classes. The last layer of the ResNet18 model is changed to fit the blockage

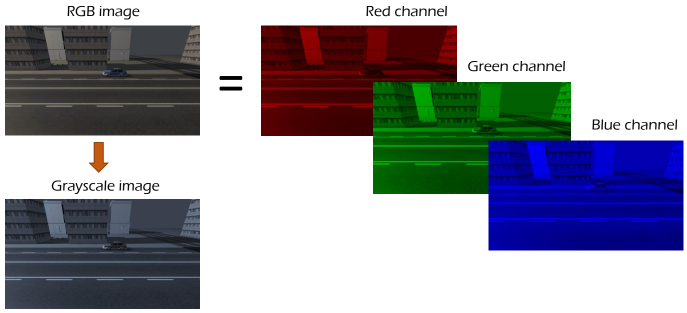


Figure 2: The RGB image is represented by red, green, and blue channels, as well as grayscale.

problem.

B. Single Channel (SICH), Image Compression (COMP) Techniques

The base work considers the RGB images of the ViWi dataset to classify them based on beam blockage status. The RGB images consist of three channels; each has several pixels (width $\times$ height) with values that reflect the image's content. Image pixels indicate the size to be stored in memory and varies according to the image resolution, which means that the higher resolution requires more storage space. Therefore, the large number of images or frames per second in videos will burden the local storage servers. On the other hand, DL algorithms outperform when the number of training datasets rises. In addition, model training is performed in cloud-centric servers that require dataset transmission through the wireless network. Accordingly, the massive amount of data transmitted over wireless links will strain the bandwidth resources of these links and cause significant delays.

Another important factor that will affect the efficiency of ML algorithms is the time required to perform model training until model convergence. Deeper DL models contain many parameters that can run into hundreds of millions or even billions, thus significantly increasing model training time, resulting in more energy consumption. In this section, our core focus is to reduce the input information and data size while maintaining the model prediction accuracy, thereby reducing the computation complexity. Precisely, as Fig. 2 demonstrates, each RGB image consists of red, green, and blue channels, and every channel has different pixel values, which means that every single image contains three folds of redundant information. Therefore, each RGB image is converted into a SICH (i.e., grayscale). After that, we shed light on reducing the storage needed to store the RGB images.

COMP is the typical technique that should be used to reduce the required storage [23]. It represents an image in fewer bits without losing the essential content information by performing data compression on digital images. COMP is generally categorised as either lossy or lossless compression. In lossless compression, the image size is reduced without affecting the quality, but the storage reduction is not significant. Therefore, our focus is directed toward lossy COMP since the main goal is to reduce the image size while maintaining image quality at an acceptable level. The most commonly used lossy COMP technique is JPEG compression. However, JPEG compression affects image quality, and a trade-off between the storage size and the image quality should be considered. Another essential factor that needs to be studied is the image quality versus the model prediction accuracy when visual information is used for model training [24]. A comparative study on the impact of SICH and COMP techniques on model performance will be presented in Section IV-B.

C. Object Detection (ODET) Technique

ODET is a CV technique that deals with recognising objects of certain classes in images or videos [25]. Since the beam blockage prediction problem in UDNs is concerned with detecting users and objects that cause signal blockage to ensure reliable communications, The ODET model can help extract such information from RGB images and exclude other insignificant details. With the aim to further alleviate the central model computation complexity and wireless transmission energy costs, this study adopts the ODET algorithm to remove the insignificant information from data. Instead of processing the entire image information, this work focuses only on image parts containing relevant information regarding the problem that the DL model is trying to solve. In this work, we exploit the pretrained you only look once (YOLO) version 3 ODET model to detect the user and the object causing the blockage. YOLOv3 belongs to the YOLO family, which contains a series of end-to-end DL models designed to provide fast ODET. The YOLO model is modified to retain the information inside the bounding boxes and remove other unnecessary information. Fig. 3 demonstrates how the modified YOLO model is used to extract the required information from each RGB image.

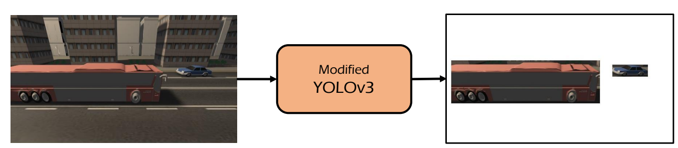


Figure 3: Using YOLO object detection to extract the required information.

IV. Performance Evaluation And Results

We evaluate the effectiveness of our proposed techniques by generating five new datasets from the original ViWi (ORIG) using simulations. Fig. 4 demonstrates these datasets, which are classified as, first, the SICH dataset that contains grayscale images. Second, the COMP dataset, which is compressed based on 20 as the quality level[1]. Third, the SICH_COMP dataset contains compressed grayscale images. The fourth dataset is generated using the ODET technique. Finally, the fifth dataset is the compressed and grayscale version of the fifth dataset (ODET_SICH_COMP).

[1]Fig. 5 demonstrates that the quality level will not affect the prediction accuracy; hence the work can use any quality level.

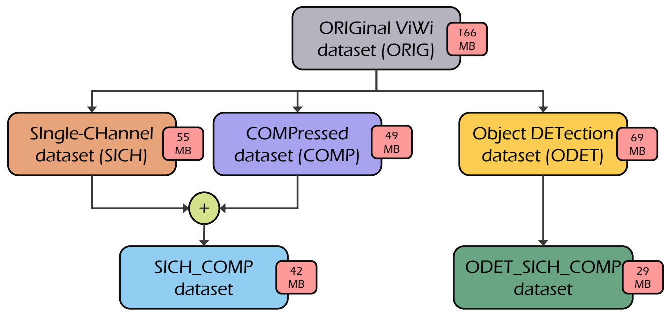


Figure 4: Different forms of ViWi dataset are generated based on our proposed techniques. Every dataset is tagged with the required memory storage size in the SBS.

A. Simulation Setup

We consider an UDN operating at a frequency of 60 GHz and contains a central server located at a macro BS and three SBSs, each with an RGB camera. In our network, the cameras capture videos, and the corresponding SBS transmit them to the central server located at the macro BS through 10Gbps mmWave backhaul links [26]. We begin our study by collecting several RGB images to train the modified ResNet18 model for the beam blockage prediction problem. We assume the cameras attached to the SBSs capture videos at 26 frames per second [27]. Therefore, to match the number of images in the ViWi dataset, which equals 5000 images, we allow each SBS to send the RGB images in 64 seconds of recording to the central server to perform initial model training. Moreover, model training hyperparameters such as number of epochs, batch size, learning rate (LR), LR schedule, LR reduction factor, data split are set to $E = 14$, $B = 150$, $LR = 1 \times 10^{-4}$, epochs 4 and 8, $\alpha = 0.1$, and 70%-30%, respectively. Our simulation experiments are based on Python programs installed on a Windows operating system with Intel Xeon CPU E5-2620 @ 2GHz and 16GB RAM. The performance of our proposed techniques is compared among the six types of ViWi datasets.

B. Simulation Results

We first investigate the effect of performing different image compression levels on the required storage and beam blockage prediction accuracy using the ORIG dataset. Fig. 5 illustrates the model prediction accuracy and the amounts of storage reduction as a function of the compression quality level. This figure reveals that regardless of the compression level used on the ORIG dataset, the prediction accuracy will remain high. Besides, it demonstrates the considerable storage reduction, which is attributed to the high compression rates that can be used. A higher compression rate means more discarded information, and hence less memory storage needed. The percentage of memory storage reduction reaches to 86% when the value 2 is used as the compression quality level.

Next, we demonstrate how this work can reduce the computational complexity of the DL model in terms of model training time while maintaining the prediction accuracy at high levels.

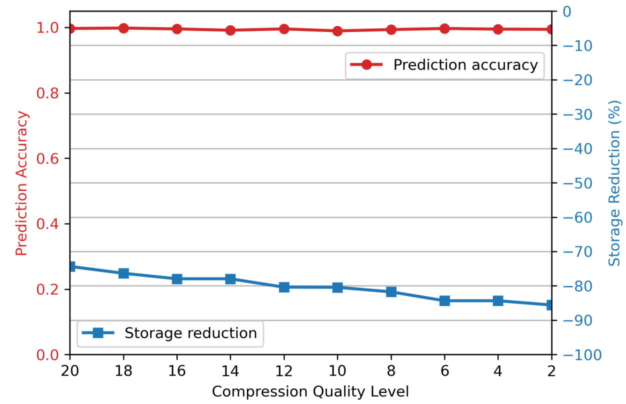


Figure 5: The effect of different compression levels on prediction accuracy and storage reduction.

Fig. 6 shows the time needed to train the ResNet18 model using the different forms of ViWi datasets. In addition, Fig. 7 illustrates the prediction accuracy achieved as a function of these datasets after 14 training epochs. It can be observed from Fig. 6 that applying JPEG COMP and ODET on the training dataset has a small impact on the training time, but the best training time reduction is achieved when we use the SICH technique. This is attributed to using single-channel images instead of three channels which means that less information needs to be processed. Therefore, fewer mathematical computations are performed within the hidden layers of the model, resulting in less time and energy consumption required to train the model until convergence. Fig. 6 also indicates that a 34% training time reduction is achieved when we use the hybrid technique that manipulates the ORIG dataset using ODET, SICH, and COMP techniques. On the other hand, Fig. 7 reveals that our proposed techniques do not affect the final prediction accuracy since all the generated datasets were able to achieve the optimal prediction accuracy.

After that, we evaluate the performance of our proposed work in the context of energy efficiency and network carbon footprint. Here, we use an average energy consumption rate of 0.0075 kWh/GB for network data transmission in 2021. This value is estimated given that the average energy intensity in the UK in 2015 was 0.06 kWh/GB, and following the same finding that the electricity intensity of data transmission is halved every two years [28]. Also, we quantify the amounts of CO_2e emissions using a conversion factor 0.23314 per kWh [28]. Fig. 8(a) shows the energy required to transmit the different types of ViWi datasets between each SBS and the macro BS. This figure reveals that all the proposed techniques significantly reduce the amount of energy consumption, and the combined data manipulation approach of ODET, SICH, and COMP is the most energy efficient achieving 82.5% reduction in energy consumption. Similarly, Fig. 8(b) demonstrates the amount of CO_2e emissions accompanied with data transmission over the wireless network. In this figure, we note that the hybrid approach has the best performance in the context of reducing the amounts of CO_2e emissions that also

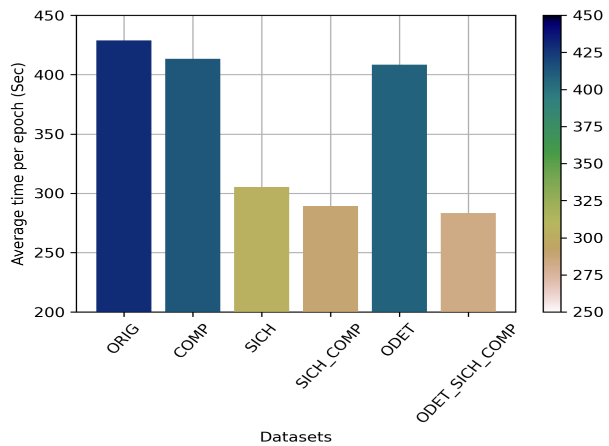


Figure 6: Average training time per epoch vs types of ViWi dataset.

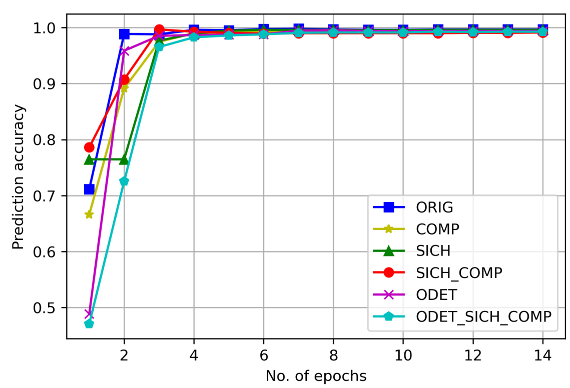


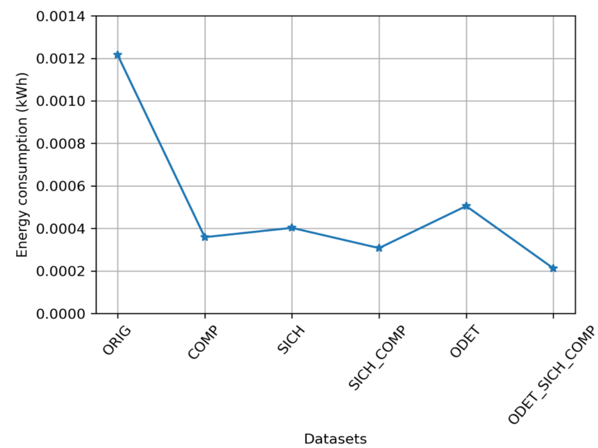
Figure 7: Blockage prediction accuracy vs number of epochs for six types of ViWi dataset.

achieves an improvement of 82.5%.

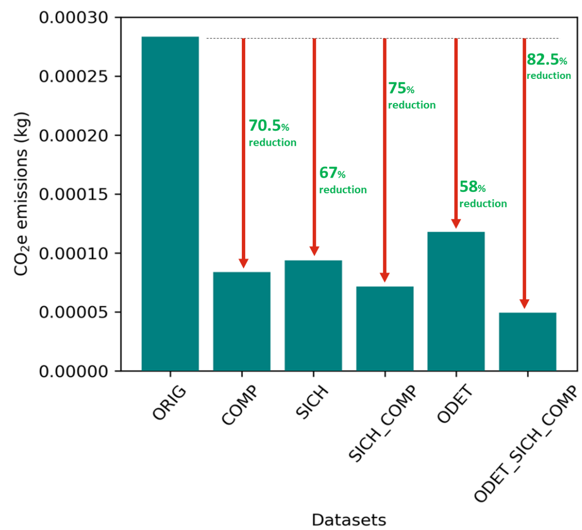
Finally, the time required to transmit the RGB images through the mmWave backhaul is plotted versus the different types of ViWi datasets in Fig. 9. It is observed that our proposed hybrid technique has significantly reduced the time required for data transmission. The reduction in data transmission time is attributed to the reduction in datasets size, which means less time to send them to the central server. Our hybrid data manipulation technique can reduce the network latency by 83% compared to that achieved by the ORIG dataset. The significant reduction in data transmission time demonstrates the potential of our proposed work in low-latency, time-sensitive applications.

V. CONCLUSIONS

In this paper, we proposed several data manipulation techniques that have improved the energy efficiency and transmission latency of DL-based vision-aided UDNs by reducing the computation complexity of DL models and the burden accompanied with transmitting large multimodal data. Network performance is improved due to removing redundant information from the training dataset and neglecting the insignificant information before starting model training. The simulation



(a)



(b)

Figure 8: The performance of using ViWi datasets in terms of (a) Energy consumption and (b) CO_2e emission

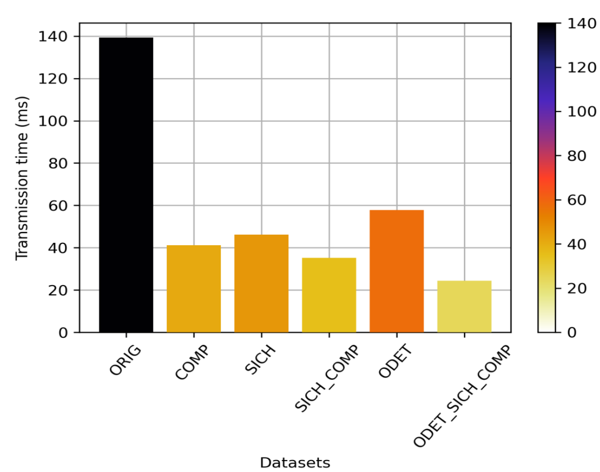


Figure 9: Transmission time needed to send the RGB images in 64 seconds of video recording from SBS to the macro BS over mmWave backhaul link.

results showed that the combination of ODET, SICH, and COMP achieved the best results by reducing the model training time by 34% and the memory size needed to store the raw data by 86% without compromising the accuracy of the DL model. Furthermore, the hybrid technique has improved the network's energy consumption and carbon footprint by 82.5% and also reduced data transmission time by 83%, which helps to realise low-latency time-sensitive applications.

References

[1] S. A. Busari *et al.*, "5G millimeter-wave mobile broadband: Performance and challenges," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 137–143, June 2018.

[2] H.-N. Dai *et al.*, "Big data analytics for large-scale wireless networks: Challenges and opportunities," *ACM Comput. Surv. (CSUR)*, vol. 52, no. 5, pp. 1–36, Sept. 2019.

[3] S. Pouyanfar *et al.*, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv. (CSUR)*, vol. 51, no. 5, pp. 1–36, Sept. 2018.

[4] Y. Zhou *et al.*, "A deep-learning-based radio resource assignment technique for 5G ultra dense networks," *IEEE Netw.*, vol. 32, no. 6, pp. 28–34, Nov. 2018.

[5] S. Moon, H. Kim, and I. Hwang, "Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications," *J. Commun. Netw.*, vol. 22, no. 3, pp. 177–184, July 2020.

[6] M. S. Sim *et al.*, "Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51 634–51 646, Mar. 2020.

[7] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *Proc. IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium*, May 2020, pp. 1–5.

[8] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC2020-Spring), Antwerp, Belgium*, May 2020, pp. 1–5.

[9] P. Monti *et al.*, "Mobile backhaul in heterogeneous network deployments: Technology options and power consumption," in *Proc. 14th Int. Conf. Transparent Opt. Netw. (ICTON), Coventry, UK*, Aug. 2012, pp. 1–7.

[10] V. Masson-Delmotte *et al.*, "Global warming of 1.5 C," *An IPCC Special Report on the impacts of global warming of 1.5*, vol. 1, no. 5, Oct. 2018.

[11] M. Denil *et al.*, "Predicting parameters in deep learning," *arXiv preprint arXiv:1306.0543*, June 2013. [Online]. Available: http://arxiv.org/abs/1306.0543

[12] Y. Guo, A. Yao, and Y. Chen, "Dynamic network surgery for efficient DNNs," *arXiv preprint arXiv:1608.04493*, Aug. 2016. [Online]. Available: http://arxiv.org/abs/1608.04493

[13] M. Rastegari *et al.*, "Xnor-net: Imagenet classification using binary convolutional neural networks," in *Proc. European conference on computer vision*, Oct. 2016, pp. 525–542.

[14] M. Mathieu, M. Henaff, and Y. LeCun, "Fast training of convolutional networks through FFTs," *arXiv preprint arXiv:1312.5851*, Dec. 2013. [Online]. Available: http://arxiv.org/abs/1312.5851

[15] M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," *Electron.*, vol. 8, no. 3, p. 292, Mar. 2019.

[16] H. Okamoto *et al.*, "Machine-learning-based throughput estimation using images for mmWave communications," in *Proc. IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, Australia*, June 2017, pp. 1–6.

[17] A. Klautau, N. González-Prelcic, and R. W. Heath, "Lidar data for deep learning-based mmWave beam-selection," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 909–912, Feb. 2019.

[18] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided dynamic blockage prediction for 6G wireless communication networks," in *Proc. IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada*, June 2021, pp. 1–6.

[19] Y. Koda *et al.*, "Handover management for mmWave networks with proactive performance prediction using camera images and deep reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 802–816, Dec. 2019.

[20] D. Amodei and D. Hernandez, "AI and compute." [Online]. Available: http://blog.openai.com/ai-and-compute

[21] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2595–2621, June 2018.

[22] M. Alrabeiah *et al.*, "ViWi: A deep learning dataset framework for vision-aided wireless communications," in *Proc. IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium*, May 2020, pp. 1–5.

[23] G. Vijayvargiya, S. Silakari, and R. Pandey, "A survey: various techniques of image compression," *arXiv preprint arXiv:1311.6877*, Nov. 2013. [Online]. Available: http://arxiv.org/abs/1311.6877

[24] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Proc. IEEE eighth international conference on quality of multimedia experience (QoMEX), Lisbon, Portugal*, June 2016, pp. 1–6.

[25] Z.-Q. Zhao *et al.*, "Object detection with deep learning: A review," *IEEE trans. neural netw. learn. syst.*, vol. 30, no. 11, pp. 3212–3232, Jan. 2019.

[26] M. M. Ahamed and S. Faruque, "5G backhaul: requirements, challenges, and emerging technologies," *Broadband Communications Networks: Recent Advances and Lessons from Practice*, vol. 43, Nov. 2018.

[27] R. Nawaratne *et al.*, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," *IEEE Trans. Ind. Inform.*, vol. 16, no. 1, pp. 393–402, Aug. 2020.

[28] S. Blenkinsop *et al.*, "Carbon emission savings and short-term health care impacts from telemedicine: An evaluation in epilepsy," *Epilepsia*, Aug. 2021.