Toward On-Device AI and Blockchain for 6G-Enabled Agricultural Supply Chain Management

Muhammad Zawish, Nouman Ashraf, Rafay Iqbal Ansari, Steven Davy, Hassaan Khaliq Qureshi, Nauman Aslam, and Syed Ali Hassan

Abstract

6G envisions artificial intelligence (AI) powered solutions for enhancing the quality of service (QoS) in the network and to ensure optimal utilization of resources. In this work, we propose an architecture based on the combination of unmanned aerial vehicles (UAVs), AI, and blockchain for agricultural supply chain management with the purpose of ensuring traceability and transparency, tracking inventories, and contracts. We propose a solution to facilitate on-device AI by generating a roadmap of models with various resource-accuracy trade-offs. A fully convolutional neural network (FCN) model is used for biomass estimation through images captured by the UAV. Instead of a single compressed FCN model for deployment on UAVs, we motivate the idea of iterative pruning to provide multiple task-specific models with various complexities and accuracy. To alleviate the impact of flight failure in a 6G-enabled dynamic UAV network, the proposed model selection strategy will assist UAVs to update the model based on the runtime resource requirements.

Introduction

The development of 6G communication has brought about several new technologies to enhance user experience by providing seamless connectivity and high quality of service (QoS). The technologies that have undergone development under the aegis of 5G networks include machine-to-machine (M2M) communication, the Internet of Everything (IoE), and device-to-device (D2D) networks, to name a few [1]. Recently, the utilization of unmanned aerial vehicles (UAVs) as access points has been explored as a viable candidate for providing on-demand services [2] for 6G networks. UAV base stations can provide connectivity to ground users in an ad hoc manner, thereby supporting the conventional network infrastructure. Moreover, artificial intelligence (AI)-empowered solutions have been proposed to enhance the QoS in 6G networks [3]. To this end, a significant amount of work has focused on using AI to optimize 6G-enabled wireless systems, but limited attention has been paid to device-level resource optimization of AI models, specifically for use in UAVs. Thus, the AI-led 6G networks will leverage the opportunities offered by the aforementioned technologies for the development of new applications.

In this work, we propose an architecture based on UAVs empowered by AI and blockchain for agriculture supply chain management by capturing images of the fields and estimating biomass through UAVs. We propose a system model based on a combination of blockchain and on-device AI for ensuring traceability and transparency, tracking the provenance of crops by observing the state of farms, inventories. and contracts in the agriculture supply chain. Typically, a UAV is only responsible for collecting data via its sensors and transmitting it to a remote cloud server for processing. Although this cloud-centric approach helps UAVs save energy by transmitting computation-intensive tasks, the wireless transmission of raw data introduces significant latency and security issues. However, shifting AI from the cloud to the device restricts unnecessary data transfer and ensures network security as a key performance indicator (KPI) of 6G-enabled wireless systems [2–4]. The proposed architecture is sensitive to both latency and privacy, which cannot be the case in a cloud-centric approach; thus, on-device computation is the preferred approach. However, UAVs usually possess limited computational and low-power capabilities, which hamper the task of on-device processing. Therefore, the computational complexity of the underlying AI algorithm plays a key role in realizing on-device processing [5].

Deep learning as an extension of AI has resulted in numerous prominent algorithms such as convolutional neural networks (CNNs) for image classification, segmentation, and so on. Recently, CNNs have caught the spotlight in a wide range of mobile vision applications including self-driving cars, cyber-physical systems, autonomous systems, and many more [5, 6]. In this work, we propose the use of a fully convolutional neural network (FCN) a type of CNN [6] for semantic segmentation of biomass pixels from an image captured by a UAV. Based on the above discussion, embedding an FCN directly into a UAV is not a feasible solution due to its limited resources. To enable on-device processing using CNNs, recent works have focused on providing a fixed compressed model using techniques such as network pruning, weight quantization, and knowledge distillation [7]. Instead of providing a single compressed model using the above techniques for on-device execution, we leverage the idea of iterative pruning to provide multiple task-specific models with various resource-accuracy trade-offs. This approach is more significant when performing on-device processing because different UAVs possess different battery timing, processing, and storage capacities [2]; thus, it is not feasible to design a model with rigid characteristics. The aim of providing models with various complexities and accracy trade-offs is to allow UAVs to fetch the required model based on dynamic resources in flight.

The motivation behind utilizing blockchain for the supply chain is to ensure transparency in information management. The system model is summarized in the flowchart shown in Fig.

Muhammad Zawish (corresponding author) and Steven Davy are with Waterford Institute of Technology, Ireland.
Nouman Ashraf is with Walton Institute, Ireland, TU Dublin, Ireland, and Turku University of Applied Sciences, Finland.
Rafay Iqbal Ansari and Nauman Aslam are with Northumbria University Newcastle, UK.
Hassaan Khaliq Qureshi and S. A. Hassan are with the National University of Sciences and Technology, Pakistan.

Digital Object Identifier: 10.1109/IOTM.006.21000112

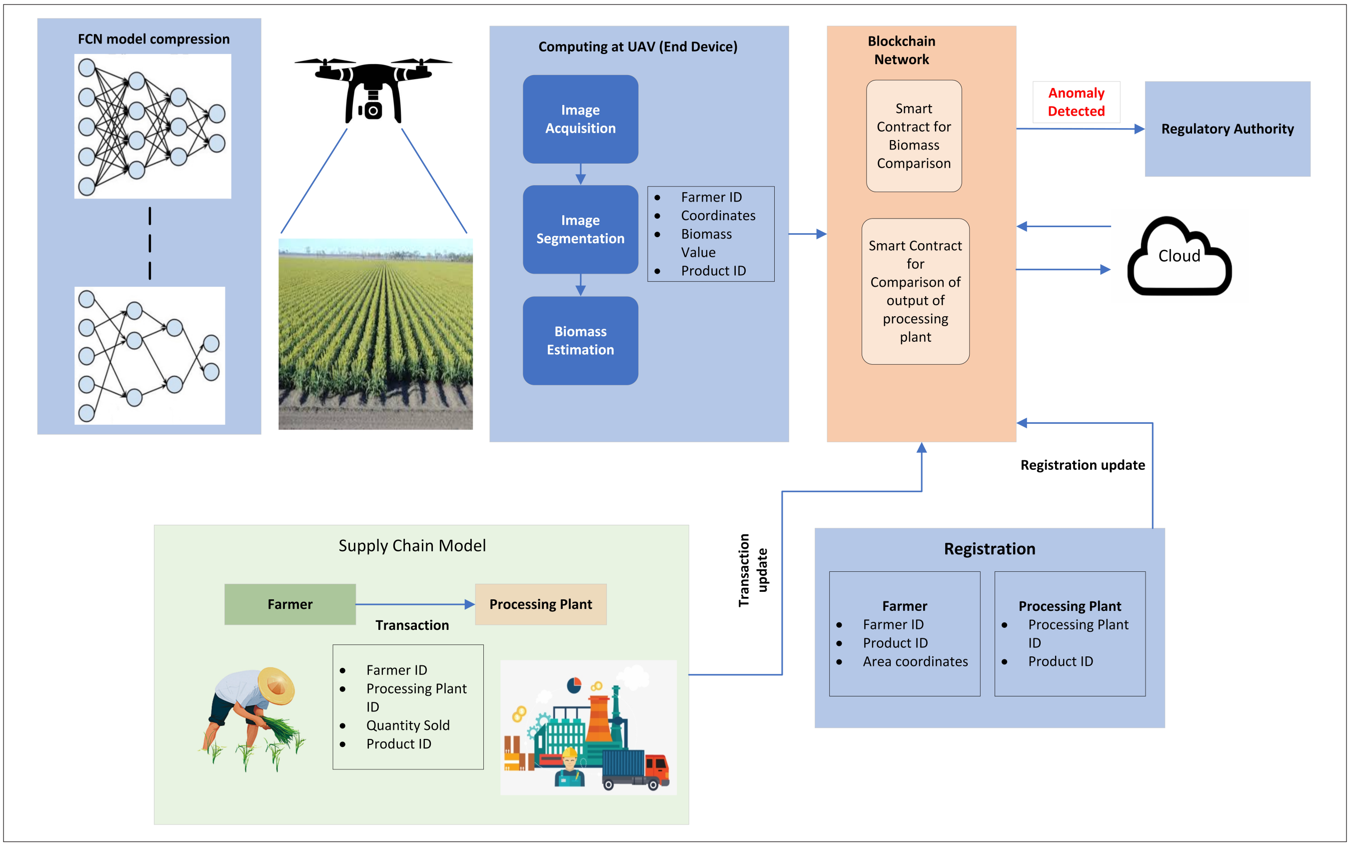

FIGURE 1. The proposed AI and blockchain based agricultural supply chain management network.

1. A detailed discussion of the flow chart is provided. The rest of the article is organized as follows. In the following section, we present the motivation behind different aspects of our model. We then present our system model, followed by the proposed approach. Following that, we present the experimental setup and results, followed by conclusions and future directions.

Motivation and Contribution

In this section, we discuss the motivation behind different aspects of the proposed model. Then we present a brief overview of the role of UAVs for 6G networks and motivate their use in agri-food supply chain management. Following that, we describe the significance of the on-device approach over the cloud-centric approach in the proposed use case. We then discuss the significance of blockchain to ensure traceability in the food supply chain. The final subsection highlights the contributions of this work.

UAVs and 6G Networks

UAVs have found wide use in several applications ranging from security to agriculture, especially in the context of providing ubiquitous connectivity for 6G networks. Global mobile traffic has seen a mushroom growth in recent years, which builds the case for exploring new technologies such as UAVs to assist the traditional networks. UAVs are particularly helpful in realizing the concept of heterogeneous networks (HetNets), where several small cells are deployed to enhance the network capacity. The main feature of UAVs that makes them an ideal solution for ad hoc networks is their agility, allowing quick and flexible deployment [8]. Hence, they can be particularly useful for emergency networks. It is envisioned that UAVs will become a part of the wider terrestrial-air integrated network. However, this integration will also lead to challenges such as allocation of resources, reliability, security, and path planning [9].

Our aim is to exploit UAVs for agri-food supply chain management by capturing and processing images of the agri fields. The main motivations behind employing a UAV for agri-food supply chain management comprises the following focal points: i) the ease of movement of UAVs makes them ideal for on-demand deployment to capture images of a field; ii) UAVs possess the ability to process and transmit information to the cloud; iii) UAVs can help cover large areas and gather images from different angles; iv) UAVs can be operated from a central location to introduce transparency for agri-food supply chain management; v) UAVs can dynamically select a lightweight AI-based solution that can reduce the bandwidth requirement and conserve resources; and vi) UAVs can provide timely updates about production to the supply chain, thereby reducing the processing costs for large volumes of centralized data.

On-Device vs. Cloud-Centric AI

Deep neural networks (DNNs) as a subclass of AI algorithms have evolved over the past several years due to availability and accessibility of:
- Big data
- Hardware accelerators
- Open source software platforms to train and optimize them [10]

The superior performance of DNNs depends on their ability to extract high-level representations from raw data generated by Internet of Things (IoT) and several other devices. Since DNNs are able to process a large amount of data, they require huge computational power to execute the floating point operations per second (FLOPS) inside their layers. A typical DNN is made up of multiple convolutional and fully connected layers stacked on top of each other, with thousands of nodes/filters interconnected in each layer that account for the millions of parameters and gigaFLOPS (GFLOPS) [5]. Therefore, for training or inference of these DNNs, traditional IoT applications rely on the powerful capabilities of cloud data centers and high-performance computers (HPCs). Generally, the raw data generated by devices is moved to the remote cloud data center for pro-

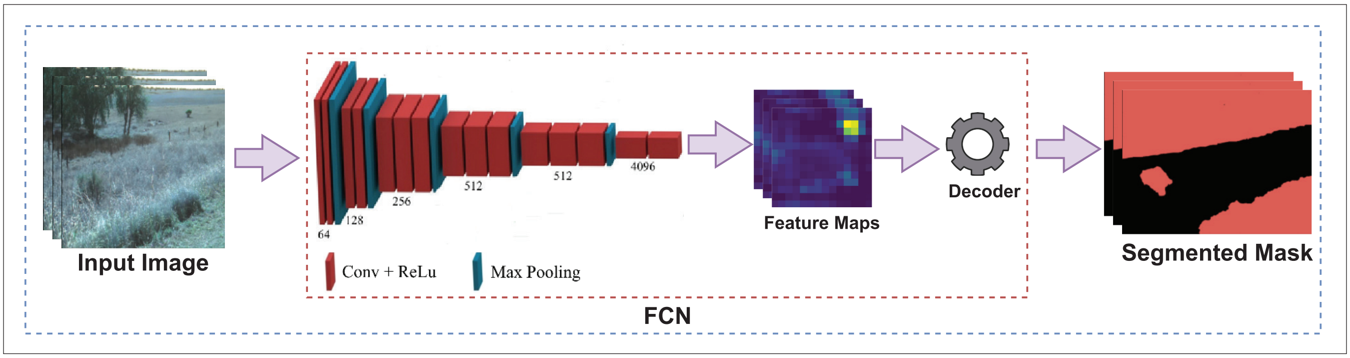

FIGURE 2. FCN architecture for semantic segmentation of biomass pixels.

cessing, and the decision is made on the cloud once inferences are made. Although this status quo approach alleviates the end devices' load, energy consumption, end-to-end latency, and privacy issues arising from wireless transmission are inevitable. On the other hand, pushing computations from cloud to end device is an emerging solution to the latency, energy, and privacy bottlenecks [2]. In this approach, the DNN model is served on the end device, which means data is being generated and processed at the same physical location. For instance, as shown in Fig. 1, our approach motivates the on-device computation in smart agricultural application by computing the biomass from images on the UAV instead of sending them to the cloud. Our approach ensures privacy and low latency by restricting the raw data and computations within the boundaries of the farm. However, the low-power end devices cannot afford the computational complexity of DNNs. This has pushed researchers to compress the DNN models so that they can be embedded into low-power end devices [7].

There are various methods to make DNNs lightweight such as filter pruning, weights quantization, and knowledge distillation. Among them, filter pruning is simple, faster, and efficient as it reduces redundant filters from the convolutional layers that account for most of the FLOPS [11]. In this approach, the filters of each convolutional layer are ranked based on a certain scoring function such as average percentage of zeros or l_1-norm. These scoring functions are used to estimate the importance of filters toward the accuracy of the task, and then only the top-*m* ranked filters are kept intact for fine-tuning [7]. This single-stage pruning approach results in a pruned and fine-tuned model with lower computational complexity for deployment on end devices. However, in the proposed scenario, we cannot rely on a model with fixed complexity due to dynamic resource demands of UAVs. Thus, we motivate the scheme of providing a family of models with various resource-accuracy characteristics so that the most efficient model can be selected based on the requirements of a certain IoT application.

Blockchain

The global agri-food supply chain is a naturally dynamic structure that has evolved since the time of hunter-gatherers through subsistence agriculture. It is now a globalized environment with many moving pieces that make it much more complex. One of the biggest challenges involves a lack of cooperation among players due to individualistic mindsets and skewed views, as well as a lack of accountability, which may contribute to food security issues [12]. Such problems can be tackled by introducing blockchain technology in agriculture. From customer desire for transparency and more knowledge about the food they purchase to record-keeping and food integrity concerns, this exciting emerging technology can hold the key.

Contributions

The contributions of this work are summarized as follows:
- We propose a novel architecture for agricultural supply chain management based on UAVs empowered by AI for on-device biomass estimation of crops.
- In contrast to traditional biomass extraction techniques, we leverage the powerful capabilities of the FCN model to provide pixel-wise label maps for biomass estimation.
- We motivate the utilization of blockchain to keep track of crops' provenance and transactions (between farmers and processing plants) for ensuring traceability and transparency in the supply chain network.
- Lastly, to operate an AI algorithm reliably over a 6G-enabled dynamic UAV network, we provide a model selection approach using iterative model compression that generates multiple task-specific AI models with various complexities and accuracy trade-offs. As opposed to single lightweight model approaches, our results show trade-off among a variety of models to enable flexible AI on UAVs by selecting the desired model.

It is to be noted that our focus in this article is not to propose a combination of AI and blockchain, but to propose an architecture based on on-device AI and blockchain for the novel application under consideration to address a serious issue in the agriculture scenario.

System Model

The system model flow diagram is shown in Fig. 1. In the proposed architecture, we assume that the UAV acts as an end device and collects images of the fields. The purpose of the proposed architecture is to collect the image data using UAVs, process it, and transmit it to the cloud. The cloud collects the information provided by the UAVs, and any transactions between a farmer and a processing plant are also uploaded to the cloud. The diagram also shows two smart contracts: the first for the biomass comparison by tracking the provenance of crops, and the second to compare the amount of raw crops that have been supplied to the processing plant and the prepared product provided by the processing plant to the market. This traceability helps in mitigating the risks of black marketing of the raw material as well as the final product at both the farm and industrial levels. For any transaction between the farmer and the processing plant (e.g., if a farmer sells sugar cane to the processing plant), the information will be updated through a blockchain network at the cloud. Each transaction generates information such as the farmer (seller) ID, the processing plant (buyer) ID, the quantity sold, and the product ID. The processing at the UAV also involves the biomass comparison with updates provided by the transactions. In the case of any anomaly between the biomass estimated value and the quantity updated by the farmer-processing plant transaction, a flag is generated and sent to the regulatory authority (e.g., government agency). Moreover, the outputs of the processing plant are also updated over a blockchain network. The aim is to check if the output of the processing plant corresponds to the input. Similar to the previous case, a flag is generated in the case of any anomalies between the processing plant output and the projected output. The purpose of the proposed architecture is to introduce transparency in the supply chain management.

Proposed Approach Based on AI and Blockchain

Based on the above system model, this section demonstrates the AI-based on-device biomass estimation and the blockchain-based smart contract.

> The purpose of the proposed architecture is to introduce transparency in the supply chain management.

On-Device Biomass Estimation

In practice, the most common approach to calculate biomass is to conduct field surveys and visually observe the height of grasses using pre-defined criteria [13]. Although this process provides accurate results, it is labor intensive and time consuming as it involves human effort. On the other hand, traditional image processing approaches rely highly on domain expertise and manual feature engineering on images to estimate biomass. However, deep learning techniques learn useful features automatically from the images with the help of convolutional filters [6]. During training of CNNs, convolutional filters play a vital role in replacing the manual feature extractors. For this reason, they require minimal human intervention and provide state-of-the-art results. We approach the task of biomass estimation using semantic segmentation of biomass pixels in the images using FCN. Figure 2 shows the proposed FCN, which takes an input image of arbitrary size and performs downsampling and upsampling using convolutional and transposed convolutional layers, respectively, to make pixel-wise predictions. The encoder part (downsampling layers) produce the class activation maps (CAMs) to increase the field of view (FOV) for the decoder part over input. Unlike traditional approaches, where images captured by end devices are forwarded to the cloud for biomass estimation, we propose on-UAV computation where a UAV is responsible for both image acquisition and biomass estimation tasks. This approach not only reduces the latency and communication cost, but also improves privacy as data does not go out of the farm. Once an image is captured by a UAV, it will be given to the FCN, which will perform semantic segmentation. Given the pixel-wise class labels of the image, the biomass will be calculated as the percentage of pixels identified as biomass. Lastly, instead of sending the images to the cloud, the UAV will only update the estimated biomass to the blockchain along with the latitude and longitude of the respective farm land.

Model Compression: The complexity of CNNs is usually measured by the number of FLOPs and amount of memory they require for execution. Therefore, the base FCN model used for on-device biomass estimation has approximately 125 GFLOPS and consumes 513 MB of memory when deployed [5]. This thus generates the corresponding need for memory requirements. Although modern UAVs are capable of handling significant amounts of computations and storage, it is not feasible to execute the raw model on them given the energy efficiency requirements. We employ filter pruning as a technique to compress the base model by removing the redundant filters from convolutional layers at a small loss of accuracy. For each filter F, we compute its l_1-norm ($\|F\|_1$) to rank its importance in a layer. The argument is that the filters with lower l_1-norm essentially generate very small activations; thus, the corresponding filters can be pruned out. The pruning ratio determines the number of filters to be removed from each layer, and the model is fine-tuned on the rest of the filters. The purpose of choosing l_1-norm over other scoring criteria is to reduce the extra computations required during the pruning process [10].

Dynamic Model Selection: When performing on-UAV inference, the key challenges posed by UAVs includes processing the incoming data with minimal delay to make critical decisions, and consuming low energy in order to navigate smoothly and maximize flight time [2]. Thus, to overcome these challenges, we propose a dynamic model selection strategy using iterative pruning. It is obvious from the discussion of model compression that accuracy can be traded off with lower complexity of compressed models. Therefore, our approach provides a roadmap of models with different accuracy, complexity, and energy trade-offs. The required model can be fetched on the fly given the dynamic resource and accuracy requirements. For instance, if the processing capacity of a UAV is depleted during a flight, the existing model can be replaced by a model having lower computational cost to avoid any failure.

Smart Contract: Blockchain

The main aim of employing blockchain in the proposed framework is to ensure transparency in the supply chain network comprising farmers and processing plants. Moreover, regulatory authorities perform audits more often, and trusting third-party auditors is not a viable solution. Therefore, we ensure verifiable auditability with the help of blockchain networks. First, farmers and processing plants are registered on the blockchain network and are assigned a unique blockchain address. This blockchain address can be used to see the historical transactions made on the blockchain network. Farmers are registered along with their unique ID, crop type, and the latitude and longitude of their crop field. Similarly, processing plants are registered along with the unique IDs and product type. Once they are registered, they are assigned a unique blockchain address, and they record all their trades on the blockchain network. Recording all the data will be advantageous for farmers as they will get good return for their crops. Moreover, the availability of different processing plants will provide farmers with the option to select one with better rates. Similarly, processing plants can easily locate farmers with different quantities and types of crops.

Once the entities are registered, they will start making transactions on the blockchain network regarding their trades. In our proposed blockchain-based architecture, we use two smart contracts: one to ensure the traceability and auditability of farmers and another for the processing plant. UAVs over the crop fields are going to assist the smart contract for biomass comparison. UAVs will capture and analyze pictures and store this information on the blockchain network. This information will help smart contract to make a comparison if the farmers harvested crops. Moreover, if they are harvested, did farmers record the trade on the blockchain network? UAV-assisted images will ensure transparency of trading. The smart contract for biomass comparison will fetch the previously stored information on the blockchain network and compare it to an image recently captured by the UAV. If the biomass data is different, smart contracts will emit an anomaly notification to the regulatory authority. Thus, the smart contract functionality will ensure transparency as well as auditability of farmers.

Similarly, we propose another smart contract in order to ensure auditability of processing plants, that is, if they are utilizing all their crops to make products. Processing plants will report their products along with their quantity on the blockchain network. This stored information will be compared to the input crops quantity recorded by the farmers. If the information shows any discrepancy, a notification alert will be emitted to the regulatory authority. Therefore, the blockchain network will ensure the auditability of the processing plants and enforce them to do their duty honestly.

The updating mechanism inside smart-contract simply involves updation of the estimated biomass value. Smart-contract involves limited number of mathematical computations and the implementation complexity is thus reasonable. However, it would induce an additional delay, which we intend to investigate in our future work. Moreover, it must be noted that

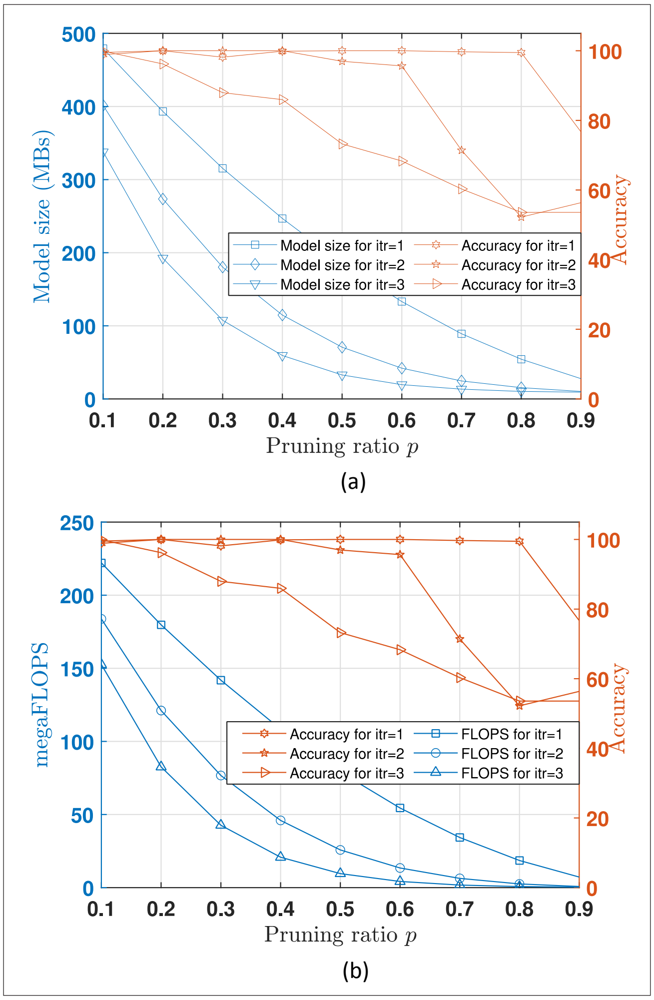

FIGURE 3. a) Trade-off between accuracy and model size over different pruning ratios p and iterations (itr); b) trade-off between accuracy and FLOPS over different pruning ratios p and iterations (itr).

Approach	Accuracy ↓	FLOPs ↓	Memory ↓
Mixed pruning [14]	5.09%	31.21%	28.89%
Network slimming [15]	8.81%	31.47%	29.53%
Proposed approach	6.18%	31.28%	29.48%

TABLE 1. Comparison with state-of-the-art approaches on FCN model. ↓ denotes the reduction in percentage with respect to the original model.

the complexity is also affected by the sampling size of the field area. The smaller the sampling size is, the higher the complexity as the computations have to be repeated more frequently and be updated via blockchain.

Experiments and Results

In this section, we provide an overview of our experiments and analyze the results achieved from iterative pruning. We show how iterative pruning can assist UAVs to adaptively select a DNN model for on-device semantic segmentation. The aim of our experiments is to explore multiple variants of a single application-specific DNN model with different computational complexities so that a suitable model can be picked up.

Training Baseline

In order to run through the pruning experiments, we initially train a baseline DNN semantic segmentation model, which is FCN. For training the FCN, we use pre-trained weights of the VGG-16 model on the ImageNet dataset to initialize the parameters of the encoder part of the FCN, while the parameters of the decoder part are initialized randomly. The network is trained using Keras deep learning application programming interface (API) with Tensorflow as backend on Nvidia Tesla K20 GPU. We fine-tune the network on a dataset obtained from [13] after performing data augmentation techniques such as scaling, contrast transformation, mirroring, and rotation. On setting a learning rate of 10^{-2} with the SGD optimizer, the model minimized the pixel-wise loss to a sufficient level on 128 epochs.

Iterative Pruning

The base model is not suitable for deployment directly on the UAV due to resource and computational constraints. On the other hand, deploying a predetermined compressed model on the device does not satisfy the dynamic requirements of UAVs. Therefore, our work addresses this challenge by using iterative pruning on a pre-trained base model to create a roadmap of descendant models with various complexity-accuracy trade-offs as illustrated in Figs. 3a and 3b. Instead of one stage, a single pruning technique where the model is pruned once with a very high pruning rate and then retrained, we use an iterative approach, which is more effective and fast. In other words, if 30 percent reduction in computational complexity is required, it is better to prune the model up to some iterations without retraining instead of pruning 30 percent at once and then retraining the model. For this reason, we have explored the model with various iterations and pruning ratios so that the probability of choosing the right model increases. We perform experiments by pruning the model up to iterations (itr) = 3, and for each iteration we vary the pruning ratio p from 0.1 to 0.9, which gives us in total 27 variants of the base model. It is interesting to note that we achieved 98 percent reduction in model size (Fig. 3a) and 99 percent reduction in FLOPs (Fig. 3b) on itr = 3 and p = 0.9 but at the cost of losing 46 percent accuracy. This scale of reduction is momentous when accuracy is not the concern for the application, but low latency and minimal energy consumption are required. However, in our proposed framework, we not only have to take care of resource consumption but also accuracy so that given the threshold of accuracy and computational complexity, the UAV can fetch the appropriate model from the local edge server, taking full advantage of 6G. Therefore, referring to Figs. 3a and 3b, it can be seen that with a small pruning ratio p = 0.2, the accuracy loss is very minimal, up to 3 percent at itr = 3, as compared to single iteration itr = 1. On the other hand, we achieve 54 percent reduction in FLOPS and 51 percent reduction in model size at itr = 3 and p = 0.2 as compared to itr = 1 and p = 0.2. Thus, iterative pruning on the same ratio p produces a much smaller model with minimal accuracy loss as compared to pruning only once. Moreover, there is no need to retrain the model because the UAV needs to replace the model as quickly as possible in order to minimize the effect on flight time and battery. We also analyze the feasibility of the model in terms of the energy consumption and inference time as shown in Figs. 4a and 4b, respectively. Inference time is measured as a delay in output required by the model to produce segmentation output. The energy consumption of the model is based on the arithmetic operations and data access on the device. We assume that each 32-bit FLOP consumes 2.3 pJ, so the energy required by arithmetic operations can be calculated by the product of FLOP count and energy consumption of a single FLOP. Regarding data access, we assume that retrieving 1MB of data from DRAM consumes 640 pJ, so for each model, it can be calculated as a product of model size and energy required to access each MB (i.e., 640 pJ). The results shown in Figs. 4a and 4b are consistent with Figs. 3a and 3b because energy efficiency and inference time are essentially dependent on model size and FLOP counts [11].

Thus, to efficiently manage the battery power and flight time of a UAV, these two parameters can also be used as selection criteria. Lastly, generating models with flexible resource-accuracy trade-offs at runtime assists mobile vision systems like UAVs to tackle dynamic resource needs by fetching the desired model.

Comparative Analysis: For the sake of comparison, we have performed comparative analysis by reproducing a few state-of-the-art approaches on model pruning. We used the same simulation parameters as mentioned in the original papers. In Table 1, we show the drop in accuracy, FLOPs, and memory after pruning FCN model using mixed pruning [14], network slimming [15], and the proposed approach. Our approach loses 6.18 percent of accuracy, which is less than network slimming, which dropped 8.81 percent. In contrast, mixed pruning loses relatively lower accuracy as compared to ours; however, this approach is based on a multi-stage process that introduces additional complexity in the compression scheme. Since our approach simply evaluates the l_1-norm of the filters, which does not put additional burden, it outweighs the drop in accuracy as compared to mixed pruning.

Usability in the UAV Scenario: We also evaluate the performance of iterative pruning on the FCN model for usability in the UAV scenario. Since UAVs are usually resource-constrained, we evaluate our proposed approach on a resource-constrained device for the sake of simulations. We measure the latency of the FCN model on an OpenStack virtual machine (VM) with 2 CPUs, 4 GB RAM, and a 10 GB hard disk. We execute the FCN model on 100 random images, and show the average latency in Fig. 5. This latency indicates the delay incurred by the FCN model in making an inference on an image while executing on a resource-constrained device. The unpruned model had approximately 107 ms of latency, while using the proposed pruning approach, the delay can be minimized with different pruning ratios and iterations, as shown in Fig. 5. This is significant for delay-sensitive applications; for instance, using pruning ratio of 0.5 over a few iterations, we can achieve approximately 50 percent reduction in latency.

Remark: It is to be noted that for the sake of this article, we have not considered the effects of mobility, which in reality will affect the battery power and communication link of the UAV with the cloud. We aim to consider these effects in our future work.

Conclusion and Future Works

In this work, we motivate the utilization of UAVs for agriculture supply chain management by capturing images of the fields and estimating biomass through UAVs. Our proposed system model is based on the combination of blockchain and on-device AI. The aim of the proposed model is to ensure transparency and traceability. We propose on-device processing of an FCN for semantic segmentation of biomass pixels from the images captured by the UAV. Moreover, to deal with the dynamic resource management on UAVs, we provide an iterative compression strategy that generates a roadmap of models with resource-accuracy trade-offs. Instead of a single compressed model, which could cause failure during a UAV flight, we show in our results that with multiple iterations over different pruning ratios, we can get multiple models for a specific task. Thereby, UAVs can update the models with respect to changing resource requirements during flight in order to be fault-tolerant.

In this paragraph, we highlight the limitations of our proposed architecture and provide a few future research directions to improve the proposed architecture. First, the utilization of an intrusion prevention system with the proposed architecture is one of the future directions of this work. Intrusion prevention can help to secure the network against any unwarranted nodes (e.g., UAVs) and lead to enhanced security. Second, a privacy preservation mechanism can be utilized with the proposed architecture, which can add anonymity and persuade those users who are concerned about privacy to participate in the network. Moreover, battery power is a major constraint for sustaining a UAV-based network. Therefore, the utilization of a hybrid energy model that utilizes alternate energy sources such

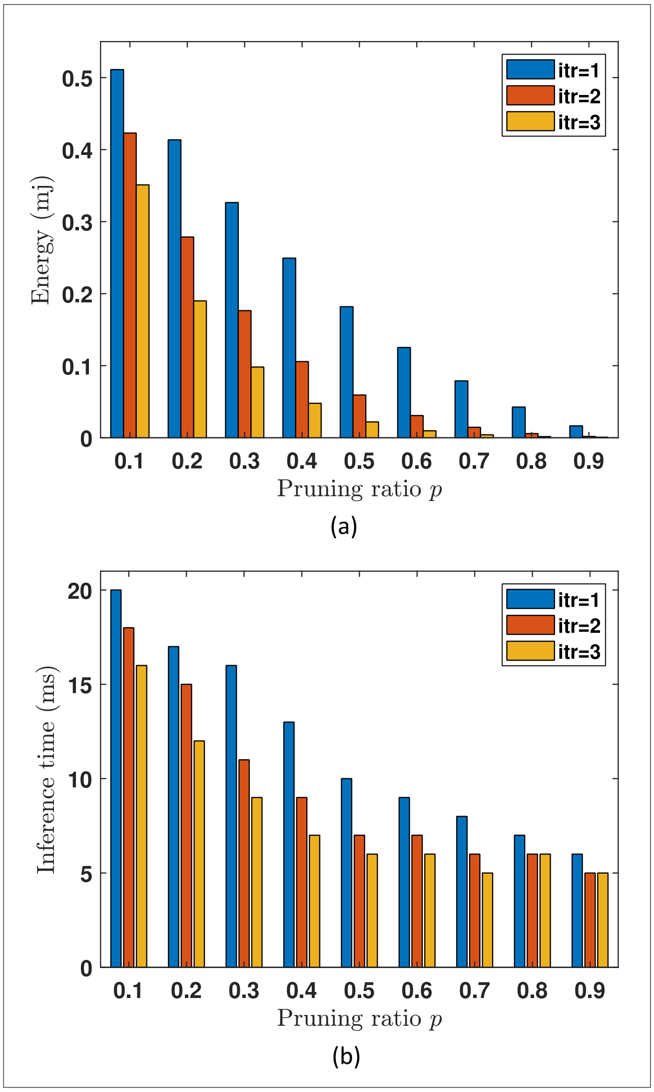

FIGURE 4. a) Depletion of energy consumption over different pruning ratios p and iterations (itr); b) depletion of inference time over different pruning ratios p and iterations (itr).

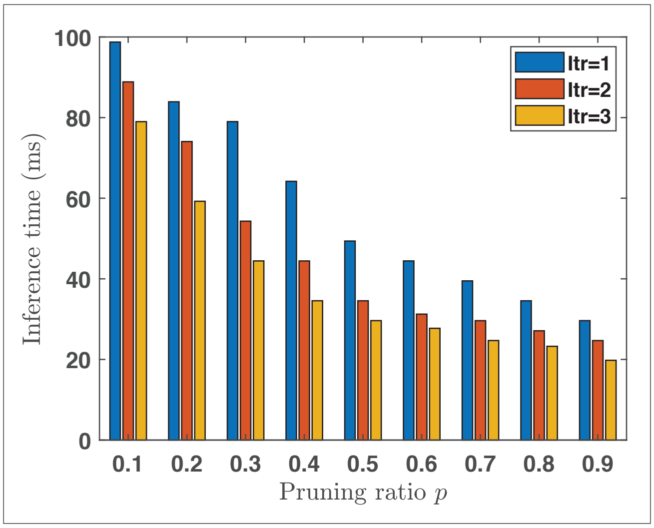

FIGURE 5. Depletion of inference time over different pruning ratios p and iterations (itr).

as solar can help enhance the network sustainability. The selection of the AI-based model can be done dynamically based on the available energy resources.

Acknowledgments

This research was supported by Science Foundation Ireland and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland VistaMilk research center under the grant 16/RC/3835.

References

[1] M. Sun *et al.*, "Noma-Based D2D-Enabled Traffic Offloading for 5G and Beyond Networks Employing Licensed and Unlicensed Access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, 2020, pp. 4109–24.

[2] H. Wang *et al.*, "Deployment Algorithms of Flying Base Stations: 5G and Beyond with UAVs," *IEEE IoT J.*, vol. 6, no. 6, 2019, pp. 10,009–27.

[3] H. Yang *et al.*, "Artificial-Intelligence-Enabled Intelligent 6G Networks," *IEEE Network*, vol. 34, no. 6, 2020, pp. 272–80.

[4] F. Tariq *et al.*, "A Speculative Study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, 2020, pp. 118–25.

[5] A. Canziani, A. Paszke, and E. Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications," ArXiv, vol. abs/1605.07678, 2016.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3431–40.

[7] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," arXiv preprint arXiv:1510.00149, 2015.

[8] B. Li, Z. Fei, and Y. Zhang, "UAV Communications for 5G and Beyond: Recent Advances and Future Trends," *IEEE IoT J.*, vol. 6, no. 2, 2019, pp. 2241–63.

[9] Y. Cao, S.-Y. Lien, and Y.-C. Liang, "Deep Reinforcement Learning for Multi-User Access Control in Non-Terrestrial Networks," *IEEE Trans. Commun.*, vol. 69, no. 3, 2021, pp. 1605–19.

[10] X. Zhang *et al.*, "Accelerating Very Deep Convolutional Networks for Classification and Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, 2015, pp. 1943–55.

[11] R. Xie *et al.*, "Energy Efficiency Enhancement for CNN-Based Deep Mobile Sensing," *IEEE Wireless Commun.*, vol. 26, no. 3, June 2019, pp. 161–67.

[12] J. Premanandh, "Horse Meat Scandal–a Wake-Up Call for Regulatory Authorities," *Food Control*, vol. 34, no. 2, 2013, pp. 568–69.

[13] L. Zhang *et al.*, "Density Weighted Connectivity of Grass Pixels in Image Frames for Biomass Estimation," *Expert Systems with Applications*, vol. 101, 2018, pp. 213–27.

[14] X. Chang *et al.*, "A Mixed-Pruning Based Framework for Embedded Convolutional Neural Network Acceleration," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 68, no. 4, 2021, pp. 1706–15.

[15] Z. Liu *et al.*, "Learning Efficient Convolutional Networks through Network Slimming," *Proc. IEEE Int'l. Conf. Computer Vision*, 2017, pp. 2736–44.

Biographies

Muhammad Zawish received a B.E. in computer systems from Mehran UET Pakistan in 2018. Later, he worked as a lab engineer in the Department of Software Engineering at PAFKIET, Pakistan. Currently, he is pursuing a Ph.D. in computer science at Walton Institute, Waterford Institute of Technology, Ireland, under the SFI VistaMilk project. His research interests are deep learning for IoT.

Nouman Ashraf received B.S., M.S., and Ph.D. degrees in electrical engineering. He has worked on the project VISORSURF (FETOPEN) at the University of Cyprus and VistaMilk at Waterford Institute of Technology. He is currently working at Turku University of Applied Science as a researcher. His research interests include the application of control theory for algorithm designing for emerging intelligent networks.

Rafay Iqbal Ansari received his Ph.D. in computer engineering from the Department of Computer Science and Engineering, Frederick University, Cyprus. Currently, he is working as a senior lecturer in the Department of Computer and Information Science, Northumbria University, United Kingdom. His research interests lie in optimal resource allocation, 5G device-to-device networks, intelligent surfaces, millimeter-wave networks, and public safety networks.

Steven Davy [M], received his B.A. Mod Hons in computer science from Trinity College Dublin in 2003 and his Ph.D. in computer science in 2008 from Waterford Institute of Technology. His research interests include distributed systems, network resource optimization and management, programmable autonomous systems, and energy-efficient artificial intelligence applied computer vision tasks. He is currently head of Programmable Autonomous Systems at Walton Institute for Information and Communication Systems Science. He has lead more than 50 research projects funded under national and EU programs.

Hassaan Khaliq Qureshi [M'16, SM'18] (hassaan.khaliq@seecs.edu.pk) received his M.Sc. degree in electrical engineering with first class honors from Blekinge Institute of Technology, Sweden, in 2006, and his Ph.D. degree in electrical engineering from City University, London, United Kingdom, in 2011. He is a recipient of the EU Erasmus Mundus staff research mobility and postdoctoral fellowship under the STRONG TIES and INTACT programs, respectively. He is currently working as an associate professor with the School of Electrical Engineering and Computer Science, NUST. His main research interests include wireless networks, network security, and cyber physical systems.

Nauman Aslam [M] is a professor in the Department of Computer and Information Science, Northumbria University. Before joining Northumbria University as a senior lecturer in 2011, he worked as an assistant professor at Dalhousie University, Canada. He received his Ph.D. in engineering mathematics from Dalhousie University in 2008. He is leading the Cyber Security and Network Systems (CyberNets) research group at Northumbria University. His research interests cover diverse but interconnected areas related to communication networks. His current research efforts are focused on addressing problems related to wireless body area networks and IoT, network security, QoS-aware communication in industrial wireless sensor networks, and application of artificial intelligence in communication networks. He has published over 100 papers in peer-reviewed journals and conferences. Dr Nauman is a member of IAENG.

Syed Ali Hassan [SM] received his M.S. degree in mathematics and Ph.D. degree in electrical engineering from Georgia Institute of Technology, Atlanta, and an M.S. degree in electrical engineering from the University of Stuttgart, Germany. His broader area of research is signal processing for communications. He was a research associate with Cisco Systems, Inc., San Jose, California. He is currently an associate professor with the School of Electrical Engineering and Computer Science, NUST, where he is also the director of the Information Processing and Transmission Research Group, which focuses on various aspects of theoretical communications. He has (co)authored more than 250 publications in international conferences and journals, and has organized several Special Issues/sessions as Editor/Chair in leading journals/conferences.