Omni-Dimensional Adaptation for MobileNetV3 using Bayesian Hyperparameter Tuning

Vo Long, Phan Nhat Quang, Tran Van Dat, Dang Duc Long
VN-UK Institute for Research and Executive Education - the University of Danang, Danang, Vietnam
Correspondence: Dang Duc Long, long.dang@vnuk.edu.vn
Received: 29/2/2024, revised: 20/4/2024, accepted: 3/6/2024
Digital Object Identifier: 10.32913/mic-ict-research-vn.v2024.n1.1267

Abstract: This paper proposes enhancing MobileNetV3 with Omni-Dimensional Dynamic Convolution (OD-Conv) to overcome CNNs' limitation of static convolution kernels. OD-Conv introduces multi-dimensional attention to adjust convolution kernels across spatial, input channel, output channel, and number of kernels dimensions, improving feature representation. Bayesian Optimization optimizes hyperparameters efficiently. Experiments show Omni-MobileNetV3 outperforms MobileNetV3 on CIFAR-100, Tiny ImageNet, and medical image datasets, achieving up to 3% accuracy gain while maintaining efficiency. This dynamic convolution method combined with Bayesian tuning achieves state-of-the-art results in image classification

Keywords: *Omni-dimensional convolution, MobileNetV3, Bayesian optimization*

Tiêu đề: **Sử dụng tối ưu Bayes để tối ưu siêu tham số cho mô hình Omni-Dimensional MobiletNetV3**

Tóm tắt: Bài báo này đề xuất việc cải tiến MobileNetV3 với Omni-Dimensional Dynamic Convolution (OD-Conv) nhằm khắc phục hạn chế của các mạng nơ-ron tích chập (CNN) về các kernel tích chập tĩnh. OD-Conv giới thiệu sự kết hợp đa chiều để điều chỉnh các kernel tích chập trên các chiều không gian, kênh đầu vào, kênh đầu ra và số lượng kernel, cải thiện khả năng biểu diễn đặc trưng. Phương pháp Tối ưu hóa Bayes (Bayesian Optimization) được sử dụng để tối ưu hóa các siêu tham số một cách hiệu quả. Thí nghiệm cho thấy Omni-MobileNetV3 vượt trội hơn MobileNetV3 trên các bộ dữ liệu CIFAR-100, Tiny ImageNet và hình ảnh y tế, đạt được mức tăng độ chính xác lên tới 3% trong khi vẫn duy trì hiệu quả. Phương pháp tích chập động này kết hợp với tối ưu hóa Bayes đạt được kết quả tiên tiến nhất trong phân loại hình ảnh.

Từ khóa: *Omni-dimensional convolution, MobileNetV3, tối ưu Bayes*

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have proven to be a valuable tool in various domains, including computer vision, natural language processing, and signal processing. The development of influential models, such as ResNet [1], Inception [2, 3], and MobileNet [4–6], has played a significant role in advancing the capabilities of computer vision systems.

At the core of CNNs lies the convolution layer, a fundamental operation that employs fixed, pre-defined kernels to extract features from input data. While undeniably effective, the static kernel does not perform well with the data outside of the training distribution[7, 8], which underscores the need for continuous innovation in CNN architectures.

To address this limitation, recent research has explored dynamic convolution techniques [9–11] aiming to learn the spatially-adaptive filters for each convolution operation, which improves the network's ability to adapt to complex patterns. However, existing dynamic convolution approaches often focus on specific dimensions of the kernel space, such as the spatial dimensions (height and width) or the input channel dimensions [9–11].

This paper proposes a novel approach, Omni-Dimensional Dynamic Convolution (ODConv), which introduces a multi-dimensional attention mechanism to dynamically adjust convolution kernels across all four dimensions of the kernel space:

- Spatial dimensions: Height and width of the kernel
- Input channel dimension: Number of channels in the input data

- Output channel dimension: Number of channels in the output feature map
- Number of kernels: Total number of kernels in the layer

However, optimizing such multi-dimensional dynamic convolution (ODConv) presents a significant challenge due to the vast hyperparameter space. Traditional hyperparameter tuning methods, like grid search or random search, can be inefficient and computationally expensive in such high-dimensional scenarios. This is where Bayesian optimization emerges as a crucial tool [12].

By leveraging past evaluations and incorporating prior knowledge, Bayesian optimization offers a more efficient and targeted approach to finding the optimal hyperparameter configuration for the ODConv layer. This can lead to significantly improved performance compared to traditional methods, unlocking the full potential of dynamic convolutions and paving the way for more powerful and adaptable CNN architectures. There are two main aspects of our contributions in this paper:

1) We apply ODConv to replace the ordinary convolutional block of MobileNetv3.
2) We explore the application of Bayesian optimization with ODConv for the new architecture's hyperparameter tuning.

II. RELATED WORK

1. Lightweight networks

In the realm of research, there has been a notable surge in the exploration of lightweight networks, emphasizing the reduction of complexity and parameters [1–4, 13]. Particularly, renowned examples, such as MobileNet, SqueezeNet [14], and ShuffleNet [15, 16], have garnered substantial attention, symbolizing the forefront of this research trend. SqueezeNet [14] introduces an innovative approach by advocating for the replacement of conventional 3x3 convolution kernels with 1x1 kernels in a "splicing" technique. This method strategically trims parameters within the current model architecture, resulting in a notable reduction in model size without compromising its functionality

2. Dynamic Weight Networks

Dynamic convolution techniques, like DyConv [9] and CondConv [10], are replacing static filters in CNNs. These methods dynamically adjust filters based on the input data, allowing the network to learn more complex features. Building on this, ODConv [17] uses a multi-dimensional attention mechanism to further enhance dynamic convolutions by adjusting filters across all their dimensions.

3. Hyperparameter Optimization

Hyperparameter optimization [18] is crucial for maximizing machine learning model performance. While traditional methods like random and grid search provide a starting point, they can be computationally inefficient. Advanced techniques such as Bayesian optimization [12] and evolutionary algorithms offer more sophisticated approaches for exploring the hyperparameter space. The comparison is illustrated in the Table I

Additionally, Neural Architecture Search (NAS) [19, 20] has the potential to improve model performance by automating architecture selection. However, NAS typically requires significant computational resources and may overfit certain datasets. Despite these challenges, these techniques collectively contribute to enhancing the effectiveness of machine learning models.

III. BACKGROUND

1. MobileNetV3

MobileNetV3 [6], a convolutional neural network (CNN), is tailored for mobile and embedded devices, balancing efficiency and accuracy. It builds upon MobileNetV1 [4] and MobileNetV2 [5], integrating hardware-aware architecture search (NAS) [6] for optimal performance across platforms. Notable features include Squeeze-and-Excitation [21] (SE) blocks for dynamic feature recalibration and hard-Swish activation for computational efficiency instead of ReLU6. Available in Large and Small versions, it caters to diverse use cases, prioritizing either accuracy or computational savings. MobileNetV3 facilitates robust deep-learning applications on resource-constrained devices, outlined in Figure 1, which illustrates its architecture: starting with a 1x1 convolutional layer, followed by depth-wise convolution with a 3x3 kernel size. Subsequent steps include SE excitation for enhanced feature representation, followed by a 1x1 convolutional layer to reduce parameters and computations, ultimately generating the output category.

2. Omni-dimentional Dynamic Convolution (ODConv)

Omni-dimensional Dynamic Convolution (ODConv) [17] presents a novel perspective on dynamic convolutional kernels within Convolutional Neural Networks (CNNs), intending to bolster the representation capabilities of deep CNNs. In contrast to conventional static convolutional kernels, which learn a singular kernel within each convolutional layer, ODConv introduces a multi-dimensional attention mechanism. This mechanism enables the learning of varied attentions for convolutional kernels across all four dimensions of the kernel space simultaneously. This process is depicted in the Figure 2.

Table I
COMPARISON OF HYPERPARAMETER OPTIMIZATION METHODS.

Feature	Grid Search	Random Search	Bayesian Optimization
Methods	Exhaustive	Random Sampling	Statistical Modeling
Computational Cost	High	Low	Moderate
Efficiency	Low	Moderate	High
Ease of Implementation	Easy	Easy	Moderate

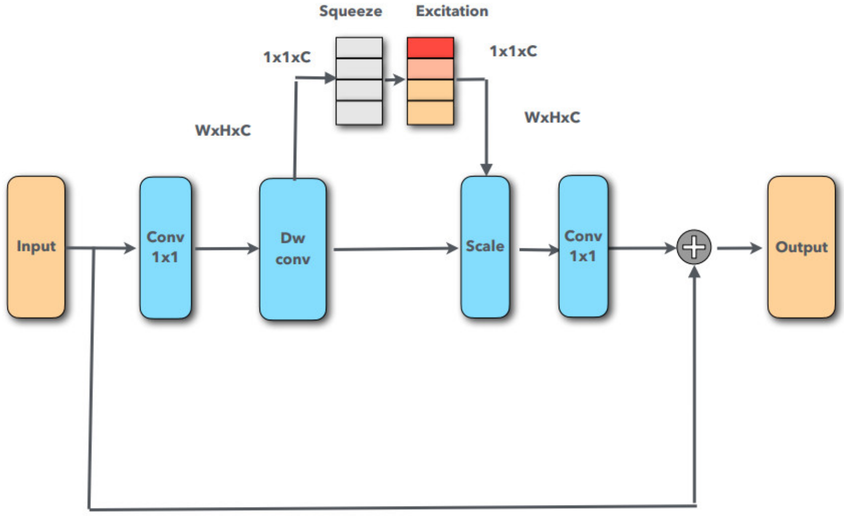

Figure 1. Structure of MobileNetV3 network model.

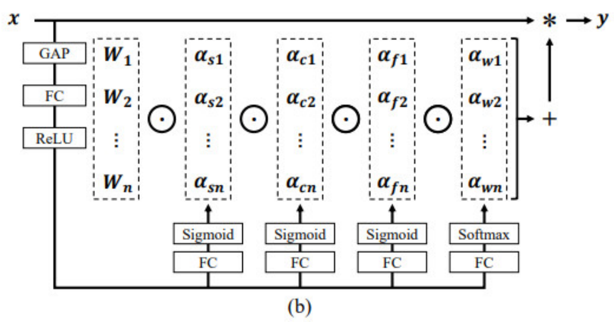

Figure 2. Schematic of Omni Dimensional Dynamic Convolution

a) Formulation for Omni-Dynamic Convolution

$$y = \left(\sum_{i=1}^{n} \alpha_{wi} \odot \ \alpha_{fi} \odot \alpha_{ci} \odot \alpha_{si} \odot \mathbf{W}_i \right) \mathbf{X} \qquad (1)$$

where $\alpha_{w_i} \in \mathbb{R}$ represents the attention scalar for the convolutional kernel $\mathbf{W}_i$; $\alpha_{s_i} \in \mathbb{R}^{k \times k}$, $\alpha_{c_i} \in \mathbb{R}^{c_{\text{in}}}$ and $\alpha_{f_i} \in \mathbb{R}^{c_{\text{out}}}$ are denoted as the attention for the spatial dimensions, the input channel dimensions, and the output channel dimensions for the kernel space of the convolution kernel $\mathbf{W}_i$ respectively. The operator $\odot$ indicates the element-wise multiplication across different dimensions of the kernel space.

b) Learning mechanism of ODConv

In ODConv [17], each convolutional kernel $\mathbf{W}_i$ is enhanced with four types of attention mechanisms (1) α_{si} allocates varying attention scalars to convolutional parameters (per filter) across $k \times k$ spatial locations. (2) α_{c_i} assigns diverse attention scalars to the c_{in} channels of each convolutional filter $\mathbf{W}_i$. (3) α_{fi} distributes distinct attention scalars to the c_{out} convolutional filters. (4) α_{wi} assigns an attention scalar to the entire convolutional kernel $\mathbf{W}_i$. The process of combining these four types of attention across n convolution is illustrated in the image above. ODConv learns four types of attention for convolutional kernels, which are shown to be complementary to each other. By progressively applying attention to the corresponding kernel, ODConv significantly strengthens the representation power of CNNs, leading to improved performance in tasks like image classification. Despite its enhanced feature learning ability, ODConv maintains efficiency by competing with or outperforming existing dynamic convolution counterparts with multiple kernels using just one single kernel. This efficiency is crucial in reducing the number of parameters and model size, making ODConv a practical and effective solution for various CNN architectures.

3. Bayesian Optimization

Bayesian Optimization [12] serves as a valuable method for optimizing objective functions that are expensive to assess, making it particularly well-suited for determining the optimal hyperparameters in deep learning tasks. It builds on a surrogate model for the objective and quantifies the uncertainty in that surrogate using the Bayesian machine learning technique, specifically the Gaussian Process, and then utilize an acquisition function defined from the surrogate model to decide where to sample in the next step.

a) Gaussian Process

Suppose there is a collection of vector points $x_1, \ldots, x_k \in \mathbb{R}^d$ and the function's values according to each point into the vector $[f(x_1), \ldots, f(x_k)]$. The vectors are drawn randomly by the prior probability distribution, The GP regression takes this distribution to be a multivariate normal with a specific mean vector and a covariance matrix.

$$f(x_{1:k}) \sim \mathcal{N}(\boldsymbol{\mu}(x_{1:k}), \boldsymbol{\Sigma}(x_{1:k}, x_{1:k})) \tag{2}$$

where $\boldsymbol{\mu}$ is the mean function that $\boldsymbol{\mu}(x_{1:k}) = [\mu(x_1), \ldots, \mu(x_k)]$ and $\boldsymbol{\Sigma}$ is the covariance matrix which:

$$\boldsymbol{\Sigma}(x_{1:k}, x_{1:k}) = \begin{bmatrix} \Sigma(x_1, x_1) & \cdots & \Sigma(x_1, x_k) \\ \vdots & \ddots & \vdots \\ \Sigma(x_k, x_1) & \cdots & \Sigma(x_k, x_k) \end{bmatrix} \tag{3}$$

b) Acquistion Functions

The acquisition function is a criterion that quantifies the utility of evaluating the objective function at a new point, given the current posterior distribution. It balances the trade-off between exploration (sampling points where the uncertainty is high) and exploitation (sampling points where the expected value is high). We choose *Expected Improvement* (EI), as a measure to identify the expected amount of improvement over the current best-observed value.

The formula for *Expected Improvement* is defined as:

$$\mathrm{EI}_n(x) = E_n[\max(f(x) - f_n^*, 0)] \tag{5}$$

The next point to sample is the one that maximizes the EI, which can be done using standard optimization methods.

$$x_{n+1} = \mathrm{argmax} EI_n(x) \tag{6}$$

IV. DATASET

In this study, we employ a diverse set of datasets to evaluate the performance of our proposed model across various computer vision tasks. The datasets used in our experiments include:

(a) (b) (c)

Figure 3. (a) Scabby apples leaves and healthy apples leaves, (b) Powdery mildew and healthy cherry leaves and (c) Rusted corn leaves and healthy corn leaves.

a) CIFAR-100 [22]

CIFAR-100 extends the complexity with 100 classes, organized into 20 superclasses. Split: The dataset consists of 50,000 training images and 10,000 test images.

b) Tiny ImageNet [23]

A subset of the extensive ImageNet dataset, Tiny ImageNet comprises 200 classes with 500 training, 50 validation, and 50 test images per class. Split: The training set contains 100,000 images, and there are 10,000 images each for validation and testing.

c) Plant Disease [24]

Focused on plant health assessment, this dataset includes 87K RGB images of leaves from three different plant species. It encompasses 38 classes relating to each state of disease of different type of species (apple, blueberry, corn, grape, orange, peach, pepper and strawberry) (Fig 3)

d) Covid-Chest Xray [25]

It has 3 different classes: Healthy chest, COVID, pneumonia. The images in this dataset have different heights and lengths so we have brought it to the same size of 448 pixels for both dimensions in each image. The total number of images in the data set is 6432 images, the training data set is 5144 images and the validation set is 1288 images (refer to Fig 4, 5, 6 for visual representation).

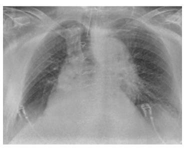

Figure 4. COVID 19

Figure 5. Normal

Figure 6. Pneumonia

V. METHODOLOGY

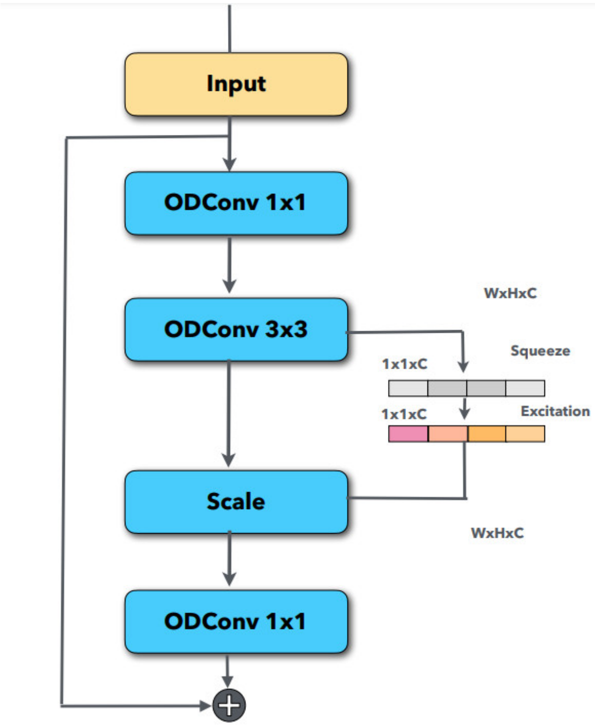

Figure 7. Structure of Omni-MobileNetV3, the combination of Omni-Dimentional Dynamic Convolution and MobileNetV3.

1. Omni-MobileNetV3 Architecture

We introduced a novel multi-dimensional attention mechanism, called "Omni-MobileNetV3", which dynamically adapted the convolutional kernels in each layer to the specific features present in the input. This allowed the network to learn more expressive filters that can better capture the relevant information for the task at hand.

We replaced the traditional convolutions in MobileNetV3 blocks with ODConv to create a hybrid architecture combining several advantages:

- **Improved Feature Extraction**: ODConv's multi-dimensional attention mechanism allowed the network to dynamically adjust convolutional kernels to the specific features in the input. This enabled the network to learn more expressive filters compared to static filters, potentially leading to improved accuracy.
- **Maintained Efficiency**: The core architecture of MobileNetV3, particularly its depthwise separable convolutions and bottleneck blocks, remained intact. This ensured that the computational efficiency and lightweight nature of the original model were preserved.
- **Increased Adaptability**: ODConv's dynamic nature granted the model an increased ability to adapt to diverse datasets and tasks. This flexibility opened doors for broader application of this hybrid architecture.

However, further exploration was needed to fully realize these benefits:

- **Balancing Complexity**: Finding the right balance between the inherent efficiency of MobileNetV3 and the added complexity of ODConv was crucial to ensure optimal performance on resource-constrained platforms.
- **Hyperparameter Tuning**: Optimizing both architectures required extensive hyperparameter tuning. This was essential to achieve the desired balance between accuracy and efficiency.

As discussed, the combination of MobileNetV3 and Omni-Dimensional Dynamic Convolution (ODConv), forming Omni-MobileNetV3, holds immense potential for efficient and accurate deep learning models. However, challenges remain in balancing complexity and finding optimal hyperparameters.

2. Bayesian Optimization

We want to enhance the performance of Omni-MobileNetV3 by fine-tuning its hyperparameters. Compared to methods like Grid Search and Random Search, Bayesian Optimization finds the optimal solution more quickly. This efficiency is due to its explore-exploit strategy, which is particularly advantageous for expensive evaluations. We focus on two hyperparameters: the number of attention kernels and the softmax temperature in the ODConv layers of each Inverted Residual block. We utilize Bayesian Optimization, which proves to be an efficient method that can search a large and complex hyperparameter space, to find the best setting for each block. The hyperparameter space of the number of kernels is from 1 to 4, and the range of softmax temperature is from 10 to 60. We begin with 5 instances generated by the Sobel Random Generator and then use the Gaussian Process and Expected Improvement algorithm to generate new combinations, from 10 to 15 instances. We evaluate the accuracy of each instance after 60 epochs of training.

VI. RESULT AND DISCUSSION

The detailed configuration of the Omni-MobileNetV3 model after applying Bayesian Optimization is in the Table II. The optimization resulted in the use of 4 3×3 kernels, 2 1×1 kernels, a temperature setting of 50, and 3 OD blocks. Note that, InvertedResidualOD is the ordinary Inverted Residual block used in MobileNetV3 but the traditional convolution layers are replaced with Omni-Convolution layers. We use the same structure as MobileNetV3-Small.

In this study, we employ the OmniMobileNetV3 model, training it for 100 epochs with optimized hyperparameters.

Table II
NUMBER OF PARAMETERS IN EACH FEATURE BLOCK OF THE MOBILENETV3 MODEL.

Block Index	Layer Type	Parameters
1	Convolution 3x3	464
2	InvertedResidualOD	3079
3	InvertedResidualOD	15985
4	InvertedResidualOD	20905
5	InvertedResidual	13736
6	InvertedResidual	57264
7	InvertedResidual	57264
8	InvertedResidual	21968
9	InvertedResidual	29800
10	InvertedResidual	91848
11	InvertedResidual	294096
12	InvertedResidual	294096
13	Omni-Convolution 1x1	124770
14	AvgPooling2D	-
15	Linear	590848
16	Prediction	-

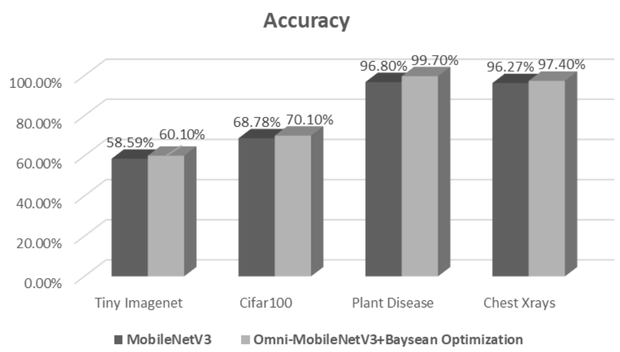

Figure 8. Compare the accuracy across four datasets

We use SGD with momentum as the optimizer, starting with a learning rate of 0.05 and momentum of 0.0004. The learning rate follows a cosine annealing schedule, and we implement temperature annealing to reduce the softmax temperature gradually for the first 50 epochs. Dataset resizing is standardized at 224×224 pixels, except for the Chest X-ray dataset, which is resized to 448×448 pixels. To improve generalization and robustness, we apply the mixup technique [26], generating synthetic data samples through linear interpolation between existing samples and their labels.

The analysis of the data table indicates that the Bayesian Optimization method led to an improvement in accuracy from 58.59% to 60.1%, a nearly 2% enhancement compared to the initial result in the Tiny-ImageNet dataset. Subsequently, we applied the identified hyperparameters from Tiny-Imagenet to the CIFAR-100 dataset to observe their impact. The accuracy increased from 68.78% to 70.1%, demonstrating a similar approximately 2% improvement. This suggests that the hyperparameters determined by the Bayesian Optimization method perform effectively across different benchmark datasets. We also extended the application of these hyperparameters to other datasets, namely Plant Diseases and Chest Xrays. Reapplying the identified hyperparameters resulted in a significant accuracy boost for the Plant Diseases dataset, rising from 96.8% to 99.7%, approximately a 3% increase. Likewise, the Chest Xrays dataset showed an accuracy improvement from 96.27% to 97.4%, roughly more than 1% compared to the original. While the accuracy improvements may look small, they are significant considering MobileNetV3 has the best performance in the class at present and is optimized carefully by NAS.

Table III
COMPARE THE PERFORMANCE OF TWO MODELS IN TINY-IMAGENET

Models	Params	Accuracy	F1-Score
MobileNetV3	**1.52M**	0.585	0.58
Omni-MobileNetV3	**1.82M**	0.596	0.59

Table IV
COMPARE THE PERFORMANCE OF TWO MODELS IN CIFAR-100

Models	Params	Accuracy	F1-Score
MobileNetV3	**1.52M**	0.687	0.69
Omni-MobileNetV3	**1.82M**	0.701	0.72
MobileNetV2	**2.36M**	0.66	0.67

Table V
COMPARE THE PERFORMANCE OF TWO MODELS IN CHEST XRAYS

Models	Params	Accuracy	F1-Score
MobileNetV3	**1.52M**	0.962	0.96
Omni-MobileNetV3	**1.82M**	0.974	0.98

Table VI
COMPARE THE PERFORMANCE OF TWO MODELS IN PLANT DISEASE

Models	Params	Accuracy	F1-Score
MobileNetV3	**1.52M**	0.968	0.96
Omni-MobileNetV3	**1.82M**	0.997	0.98

VII. ABLATION STUDY

We perform Bayesian Optimization to determine the optimal parameters for our model. Specifically, we are tuning the number of kernels in the Omni-Dimensional Dynamic Convolution with corresponding kernel sizes of 3 and 1, the temperature for flattening the attention to near-uniform to aid learning in the early epochs, and the number of OD Blocks. This search is conducted on the Tiny-ImageNet dataset over 60 epochs. The result is displayed in the (Table VII). The configuration column provides details on the number of 3×3 kernels, the number of 1×1 kernels, the temperature within an OD-Block, and the number of OD-Blocks.

Table VII
TABLE OF PARAMETERS AND PERFORMANCE METRICS

Configuration	Parameters	GFLOPS	Accuracy
2/1/29/7	1.84 M	0.26	0.5786
3/3/53/1	1.85 M	0.44	0.5819
4/2/19/4	1.84 M	0.36	0.5819
1/4/39/11	3.25 M	**0.11**	0.5835
1/2/58/9	2.04 M	0.2	0.5792
1/1/26/8	1.84 M	0.23	0.5769
3/1/33/6	1.84 M	0.28	0.5847
2/1/30/6	1.82 M	0.28	0.5776
4/2/50/3	***1.82 M***	*0.38*	***0.5867***

VIII. CONCLUSION

We proposed a novel approach integrating ODConv into the MobileNetV3 architecture and created a new hybrid model called Omni-MobileNetV3. By enabling kernel adaptation across spatial, channel, filter and kernel dimensions independently, ODConv enhances the network's ability to capture diverse feature representations tailored to the input. To efficiently optimize its hyperparameters, we employed Bayesian Optimization which leverages past evaluations to guide the search process.

Experiments were conducted on four diverse image classification datasets - CIFAR-100, Tiny ImageNet, Plant Disease, and Chest X-rays. The results demonstrated that Omni-MobileNetV3 optimized with Bayesian Optimization consistently outperformed the baseline MobileNetV3 architecture. Accuracy improvements ranged from 1-3% depending on the dataset, with the largest gains seen on the more challenging Plant Disease and Chest X-rays medical imaging datasets.

A deeper analysis of the results indicated that the multi-dimensional attention mechanism incorporated by ODConv played a key role in dynamically enhancing the representational power of the convolutional kernels. Bayesian Optimization also proved effective in identifying hyperparameter configurations that balanced model complexity and predictive performance.

Overall, the findings validate that ODConv is an effective approach for dynamically shaping convolutional kernels based on input features. When combined with Bayesian tuning, it achieves state-of-the-art accuracy on image classification tasks. Future work includes exploring its integration with other architectures and applications beyond computer vision.

ACKNOWLEDGEMENT

This research is funded by Funds for Science and Technology of the VN-UK Institute for Research and Executive Education, the University of Danang under project number VNUK-2023-Student 03

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[6] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[9] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[10] X. Jia, B. De Brabandere, T. Tuytelaars, and L. V. Gool, "Dynamic filter networks," *Advances in neural information processing systems*, vol. 29, 2016.

[11] A. Diba, V. Sharma, L. V. Gool, and R. Stiefelhagen, "Dynamonet: Dynamic action and motion network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[12] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[13] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[14] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[15] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[16] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in

Proceedings of the European conference on computer vision (ECCV), 2018, pp. 116–131.
[17] C. Li, A. Zhou, and A. Yao, "Omni-dimensional dynamic convolution," *arXiv preprint arXiv:2209.07947*, 2022.
[18] T. Yu and H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications," 2020.
[19] H. Pham, M. Guan, B. Zoph, Q. Le, and J. Dean, "Efficient neural architecture search via parameters sharing," in *International conference on machine learning*. PMLR, 2018, pp. 4095–4104.
[20] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.
[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
[22] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
[23] Y. Le and X. Yang, "Tiny imagenet visual recognition challenge," *CS 231N*, vol. 7, no. 7, p. 3, 2015.
[24] S. Bhattarai, "New plant diseases dataset," https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset/data.
[25] P. Patel, "Chest x-ray (covid-19 & pneumonia)," https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia.
[26] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

Tran Van Dat is a second year student, studying Data Science major, at VN-UK Institute for Research and Executive Education - the University of Danang, Danang. His current research interest is Computer Vision, Natural Language Process, and Foundation Models.
Email: dat.tran220407@vnuk.edu.vn

Dang Duc Long Dr. Dang Duc Long is a lecturer at the VN-UK Institute for Research & Executive Education at the University of Danang, where he has worked since 2014. He previously served as the Head of the Research and Industrial Cooperation Office at the same institute from 2014 to 2021. Dr. Long obtained his Ph.D. in Biochemistry from the University of Massachusetts Lowell in 2003, after completing his M.Sc. in Biochemistry from the same university in 1999. His research interests encompass a diverse range of fields, including computational biology, molecular diagnostics, enzymology, and circular economy.
Email: long.dang@vnuk.edu.vn

Vo Long is a third year student, studying Data Science major, at VN-UK Institute for Research and Executive Education - the University of Danang, Danang. His current research interest is Computer Vision, Natural Language Process, and Foundation Models.
Email: long.vo210404@vnuk.edu.vn

Phan Nhat Quang is a third year student, studying Data Science major, at VN-UK Institute for Research and Executive Education - the University of Danang, Danang. His current research interest is Computer Vision, Natural Language Process, and Foundation Models.
Email: quang.phan210405@vnuk.edu.vn