Attention Mechanism Assisted Deep Learning Algorithm for Multi-Category Household Waste Classification

Shoufei Han
School of Artificial Intelligence
Anhui University of Science and Technology
Huainan, China
hanshoufei@gmail.com

Xu Liu
School of Artificial Intelligence
Anhui University of Science and Technology
Huainan, China
2652722986@qq.com

Xiaojing Liu
School of Artificial Intelligence
Anhui University of Science and Technology
Huainan, China
xjydynl@126.com

Honghao Zhu
Department of Computer Science
Bengbu University
Bengbu, China
bbxyzhh@163.com

Ruyi Han
School of Mathematics
Shandong University
Jinan, China
hanruyi@mail.sdu.edu.cn

2024 International Conference on Networking, Sensing and Control (ICNSC) | 979-8-3503-6522-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICNSC62968.2024.10759911

***Abstract*—With the improvement of people's living standards, the annual global production of waste continues to rise, but the traditional household waste classification methods are burdened with a heavy task due to the wide variety of waste types and the difficulty of identifying them. Deep learning based waste image classification methods can accurately classify the waste images. In this paper, Visual Geometry Group (VGG16) is used to handle multi-category household waste classification. In order to further improve its classification accuracy, an attention mechanism is incorporated into it, and thus a VGG16 deep learning model with attention mechanism (VGG16-AM) is proposed. Experimental results show that the classification accuracy of our proposed model on the waste dataset is significantly improved to 93% compared to other deep learning algorithms.**

***Index Terms*—multi-category household waste classification, deep learning, VGG16, attention mechanism**

I. Introduction

With the acceleration of urbanisation and the improvement of people's living standards, the generation of household waste has shown a trend of rapid growth. Traditional waste disposal methods, such as landfill and incineration, not only take up a large amount of land resources, but also may cause secondary pollution to the environment. Therefore, the implementation of scientific and effective household waste classification is of great significance in improving the resource recycling rate, reducing environmental pollution and promoting sustainable development.

Household waste classification [1] involves a number of categories, including recyclables, hazardous waste, wet waste (food waste) and dry waste [2]. Traditional classification methods mainly rely on manual identification, which is not only inefficient, but also susceptible to subjective factors, resulting in less accurate classification results. In recent years, the rapid development of deep learning technology [3] provides new solution ideas for household waste classification. Deep learning algorithms [4]–[6] have powerful feature extraction and pattern recognition capabilities, and are able to automatically learn and recognise complex image data, so they are widely used in image classification [7], object detection [8] and other fields.

However, there are still some challenges in applying deep learning techniques directly to multi-category household waste classification. Firstly, there are many types of household waste, and the features between different categories vary greatly, which requires the algorithm to be able to accurately extract and distinguish these features. Second, image data of household waste often has problems such as noise, occlusion, and light changes, which can negatively affect the classification results.

To solve the above problems, this work proposes an attention mechanism-assisted deep learning algorithm for multi-category household waste classification. VGG16 [9] is selected as a representative of deep learning algorithms to show how the attention mechanism can be incorporated, and VGG16 with an attention mechanism (VGG16-AM) is designed. By introducing the attention mechanism, the algorithm is able to automatically focus on key regions in the image and ignore irrelevant background information, thus improving the accuracy and robustness of classification.

This work aims to make the following new contributions:

1) This work successfully combines the attention mechanism with deep learning algorithms to propose a VGG16-

This work was supported by Natural Science Foundation of China (62306279); the Key Projects of Natural Science Research in Anhui Higher Education Institutions (2022AH051909); Anhui province science and technology specialists in the field of agricultural material and technical equipment to unveil the list of marshal project (2022296906020001). (Corresponding authors: Xiaojing Liu)

AM model for multi-category household waste classification. By introducing the attention mechanism, the model is able to automatically learn and recognise the key regions in the image, thus improving the accuracy and efficiency of classification; and

2) This work provides a new solution for multi-category household waste classification. The superiority of the proposed algorithm is verified by comparing it with other deep learning algorithms on a large number of household waste images.

The rest of this work is organized as follows. Section II reviews the related work about the solutions to waste classification, and Section III describes the proposed algorithm in detail. Section IV analyzes the experimental results. The last section concludes this work and provides a vision for the future.

II. Related Work

A. Waste Classification

Waste classification affects people's daily life, many scholars have started to pay attention to this field and come up with different methods to tackle it.

Hossen et al. [10] designed a reliable and robust deep learning model aiming to achieve effective recyclable waste classification. Through the deep learning techniques, the model is able to automatically and accurately identify and classify different types of recyclable waste, thus improving the efficiency and accuracy of waste classification.

Belsare et al. [11] explored a method for solid waste image classification in smart cities by integrating the Internet of Things (IoT) and Wireless Sensor Networks (WSN) with wavelet transform and machine learning. Specifically, they developed an integrated system that collects and transmits waste image data through IoT and WSN, then used wavelet transform for image processing and feature extraction, and finally applied machine learning algorithm for classification.

Molina et al. [12] applied super-resolution to enhance the clarity and detail of low-resolution images from landfill sites, thereby improving the accuracy of waste classification. The results indicate that this approach significantly enhances classification precision.

Hou et al. [13] first discussed the development of a deep learning model for classifying waste plastic bottles, and then utilized advanced deep learning techniques to accurately identify and categorize different types of plastic bottles, aiming to improve recycling processes. Their results demonstrate that the deep learning model achieves high classification accuracy.

In [14], the authors presented the Real-Waste dataset, which contains real-life images of landfill waste, to facilitate the development and testing of deep learning models for waste classification. They highlighted the effectiveness of using this dataset to train deep learning algorithms, resulting in improved accuracy and robustness in classifying various types of landfill waste.

Zheng et al. [15] proposed an efficient algorithm, namely Focus-RCNet, to classify recyclable waste. It leverages focus mechanisms and knowledge distillation to enhance its performance while maintaining a lightweight structure. The results demonstrated that Focus-RCNet achieves high accuracy in waste classification, making it suitable for real-time applications in recycling systems.

B. Deep Learning with Attention Mechanism

The attention mechanism is integrated with deep learning aiming to enhance the model's performance, particularly in handling complex and long-sequence data. Thus, many scholars focus on this.

Jiang et al. [16] explored the development of fusion networks incorporating high-order attention mechanisms for improved 3D object detection in autonomous driving. Specifically, they proposed a novel deep learning architecture that integrates high-order attention to better capture complex spatial relationships and enhance feature representation. It can significantly boost the accuracy and reliability of 3D object detection in autonomous driving environments.

Liang et al. [17] presented a method for echocardiographic image segmentation using a semi-supervised deep learning approach enhanced by attention mechanisms. In it, they developed a model that leverages limited labeled data alongside a larger set of unlabeled data, employing attention mechanisms to focus on important features and improve segmentation accuracy, which significantly enhances the precision and reliability of echocardiographic segmentation.

In [18], the authors combined the genetic algorithm with deep learning and attention mechanism to optimize the predictive model. The genetic algorithm is used to fine-tune the deep learning model's parameters, while the attention mechanism helps the model focus on the most relevant features for accurate soil moisture prediction. This integrated approach can significantly improve the accuracy and efficiency of the model.

Zhang et al. [19] presented a deep learning approach for spatial downscaling of ESA CCI soil moisture data. In it, they developed a model that incorporates the attention mechanism to enhance the spatial resolution of coarse soil moisture data, which can effectively improve the accuracy the downscaled soil moisture information.

Chen et al. [20] introduced an advanced deep learning network designed for detecting ships in synthetic aperture radar (SAR) images, particularly in complex scenes. The proposed network integrates deformable convolution, which allows the model to adapt to geometric variations of ships, and attention mechanisms, which help the model focus on significant features within the images. This combination can enhance the detection accuracy and robustness.

In [21], the authors presented a multitask deep learning approach for estimating soil properties using visible and near-infrared (Vis-NIR) spectra. They incorporated attention mechanisms to enhance feature extraction and employed loss-weight balancing to optimize the training process across multiple tasks. The results show that the proposed model can significantly improve the accuracy and efficiency of soil property estimation.

C. Novelty of This Work

This work differs from previous work in that 1) it combines VGG16 and the attention mechanism, i.e., it incorporates the attention mechanism into the VGG16 network, making it more capable of focusing on the important parts of the rubbish pictures; and 2) the proposed model is used for multi-category household waste classification, which can effectively improve the accuracy of classification.

III. Proposed Model

In this section, the VGG16 and attention mechanism are first introduced, then we show how they can be combined for multi-category household waste classification.

A. VGG16

VGG16 [22] is a deep convolutional neural network proposed by the Visual Geometry Group at the University of Oxford, which performed well in the ImageNet Image Recognition Challenge in 2014.The main feature of VGG16 is that it increases the depth of the network by stacking convolutional layers with small convolutional kernels (3×3) to capture more complex features.

VGG16 consists of 16 weighted layers, including 13 convolutional layers and 3 fully connected layers, and 5 max pooling layers.

In it, the convolution layer uses a 3×3 convolution kernel for the convolution operation with the following formula:

$$Z_{i,j,k} = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{c=1}^{C} W_{m,n,c,k} \cdot X_{i+m-1,j+n-1,c} + b_k, \quad (1)$$

where $Z_{i,j,k}$ is the output of the k-th convolution kernel at position $(i,\ j)$. W is the weight matrix of the convolution kernel, X is the input feature map, b_k is the bias of the k-th convolution kernel, and C is the number of channels of the input feature map. M and N are the height and width of the convolution kernel, respectively, which are set to 3 in VGG16.

After that, the ReLU function is employed as an activation function to perform a nonlinear transformation to obtain the final output:

$$Z^{'}_{i,j,k} = ReLU(Z_{i,j,k}). \quad (2)$$

In VGG16, the maximum pooling layer uses a 2x2 pooling kernel with a step size of 2, and it is used for down-scaling and reducing the amount of computation, and the pooling operation can reduce the size of the feature map and retain the important features.

VGG16 improves the classification performance of the model by increasing the depth of the network, which can learn more complex features and patterns. The use of a small 3x3 convolutional kernel reduces the number of parameters while maintaining the size of the Receptive Field (RF). This design is both effective and efficient. Due to the above mentioned advantages, VGG16 has been noticed by scholars and used in a wide variety of fields, and thus this work uses it as a base method for multi-category household waste classification.

B. Attention Mechanism

Attention Mechanism (AM) is a technique in deep learning designed to allow models to more effectively focus on the important parts and ignore irrelevant information when processing input. Self-attention mechanism [23] is a popular one among Attention Mechanisms. In it, each position in the input sequence is subjected to a similarity calculation with all other positions in the sequence. Specifically, the steps of the self-attention mechanism are as follows:

(1) Calculate query, key and value: Given the input matrix X with the size of $(T,\ d)$, where T is the length of the sequence and d is the feature dimension. The Query, Key and Value matrices are obtained by three linear transformations:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V, \quad (3)$$

where W_Q, W_K and W_V are the query, key and value weight matrices, respectively.

(2) Calculate attention scores: The attention scores are computed using the scaled dot product formula, indicating the similarity between the query and the key, which can be calculated as:

$$e_{ij} = \frac{Q_i \cdot K_j^T}{\sqrt{d_k}}, \quad (4)$$

where d_k is the dimension of the key vector, which is used to scale the score to prevent the value from being too large.

(3) Calculate attention weights: the Softmax function is used to covert the attention scores into attention weights, which can be calculated as:

$$a_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{T} exp(e_{ik})}. \quad (5)$$

(4) Calculate the output: the attention weights are used for weighted summation of the value vectors, which in turn yields the output of the self-attention:

$$z_i = \textstyle\sum_{j=1}^{T} a_{ij} V_j, \quad (6)$$

where z_i is the output vector, the output vector of the i-th position contains the weighted information of all positions in the sequence.

C. VGG16 with Attention Mechanism

In this work, we combine the advantages of VGG16 and the attention mechanism, and propose an attention mechanism-assisted VGG16 (VGG16-AM), and its neural network structure is shown in Fig. 1.

From Fig. 1, it can be seen that the attention mechanism is added to the penultimate layer of VGG16 aiming to improve the performance of models in processing complex image tasks. First, the pre-trained VGG16 model is loaded and the top fully connected layer is removed. It allows feature extraction using its convolutional layer. Then, the penultimate four layer is found in the VGG16 model and the attention layer is inserted

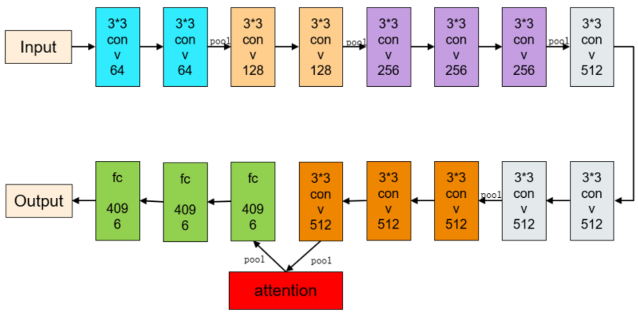

Fig. 1. Neural network architecture of VGG16-AM.

after it. After adding the attention layer, the subsequent pooling layers of the VGG16 model continue to be built aiming to keep the original structure. After the final pooling layer, a fully connected layer operation is performed to connect the modified VGG16 model and its top layer to form a complete model structure. Finally, the model is compiled using a suitable optimizer, loss function and evaluation metrics, and then the model is trained using the training multi-category household waste classification image.

IV. Experiments and Analyses

A. Parameter Settings

All experiments are conducted on a PC running the Microsoft Windows 10 operating system. The deep learning model is implemented using TensorFlow 2.1 and programmed in Python 3.9. To ensure the reliability of the results, all experiments are executed 10 times, and the final results are averaged to reduce randomness.

The model is trained using the Adam optimizer, a popular adaptive learning rate optimization algorithm in deep learning. The learning rate is set to 0.001. The training process utilizes a batch size of 32, and the models are trained for a total of 100 epochs.

B. Performance Metrics

In this work, four performance metrics, namely Cross Entropy Loss (CEL), Accuracy, Precision and Recall [24], are selected to measure the performance of the proposed model. Smaller CEL indicates better model performance and higher values of the other three performance metrics indicate better model performance.

C. Data Set

A dataset of 5679 images covering 20 common waste categories is collected and labelled, and the dataset is thoroughly cleaned and also expanded as necessary to ensure its suitability for model training and evaluation. Among them, 3982 images are used for classification training of the model and 1697 images are used to test the classification effectiveness of the model.

D. Effectiveness of Our Proposed Model

In order to validate the superiority of our proposed model in handling multi-category household waste classification task, three deep learning algorithms are selected, they are basic convolutional neural network (CNN), basic VGG16 [9] and MobileNetsV2 [25], which uses both Linear Bottleneck and Inverted Residuals mechanisms to enhance the MobileNets that reduces the number of parameters and computation by replacing traditional convolutional layers with Depthwise Convolution and Pointwise Convolution. Their parameter settings follow their corresponding references. The results are presented in Table I, where the best result are marked bold.

TABLE I
Results achieved by CNN, VGG16, MobileNetsV2 and VGG16-AM on multi-category household waste classification dataset.

Model	CEL	Accuracy	Precision	Recall
CNN	0.9745	0.6787	0.8118	0.5619
VGG16	0.6940	0.7672	0.8718	0.6694
MobileNetsV2	0.6993	0.8343	0.8474	0.8249
VGG16-AM	**0.2565**	**0.9169**	**0.9336**	**0.9034**
MobileNetsV2-AM	**0.2146**	**0.9388**	**0.9615**	**0.9219**

Fig. 2. Training and validation loss of VGG16-AM.

From Table I, it can be seen that the CNN model has a relatively high loss and the lowest precision rate, showing weak performance in basic classification tasks. Although the precision rate is relatively high, the recall rate is low, indicating that the model misses many actual positive class samples although it is able to accurately predict the positive classes. The VGG16 model shows a significant improvement in performance over the CNN, especially in the precision rate, indicating that it is able to more accurately label positive class samples. The loss function and recall are also improved, showing better adaptability on complex data. The MobileNetV2 model has the most balanced performance among all models, with high accuracy and recall, although its precision is slightly lower than that of VGG16. The VGG model with the attention mechanism (VGG16-AM) shows excellent performance on all metrics, with a significant reduction in the loss function, while achieving the highest values in accuracy, precision and recall. This indicates that the attention mechanism effectively enhances the feature recognition and classification ability of the model, especially in accurately recognising and covering positive classes.

In addition, in order to further verify that the attention mechanism can assist deep learning to handle multi-category waste classification, we fuse MobileNetsV2 and AM, namely MobileNetsV2-AM, and the results are displayed in the last row of Table I. From it, we can see that MobileNetsV2-AM can achieve better performance in terms of CEL, Accuracy, Precision and Recall. It means that the attention mechanism can help the MobileNetsV2 well.

The training and validation loss of our proposed model is shown in Fig. 2. We can see that the loss is decreasing as the epoch increases, indicating that the model is getting better and better.

Based on the above analysis, we can conclude that our proposed model can solve the multi-category household waste classification task well, and is superior in several performance metrics compared to other deep learning algorithms.

V. Conclusion

In this work, an improved VGG16 deep learning algorithm with the attention mechanism is proposed to handle the multi-category household waste classification task. The experimental results have well provided compelling evidence to show the outstanding performance of the proposed model. Even though the proposed algorithm is able to achieve satisfactory performance, it suffers from a number of limitations: 1) the model is relatively homogeneous, and 2) the network structure is fixed.

In the future, it is worthwhile to apply other technologies to improve the performance of VGG16 so as to better handle multi-category household waste classification task.

References

[1] J. P. Q. Tomas, M. L. Guimong, G. A. Leria, and I. L. Morelos, "Multiple garbage detection using mask R-CNN in tensorflow for household waste classification," in *Proceedings of the 2023 6th International Conference on Computational Intelligence and Intelligent Systems, CIIS 2023, Tokyo, Japan, November 25-27, 2023*. ACM, 2023, pp. 147–153.

[2] A. Namoun, B. R. Hussein, A. Tufail, A. Alrehaili, T. A. Syed, and O. Benrhouma, "An ensemble learning based classification approach for the prediction of household solid waste generation," *Sensors*, vol. 22, no. 9, p. 3506, 2022.

[3] P. Zhang, W. Huang, Y. Chen, and M. Zhou, "Predicting quality of services based on a two-stream deep learning model with user and service graphs," *IEEE Trans. Serv. Comput.*, vol. 16, no. 6, pp. 4060–4072, 2023.

[4] Z. Tan, J. Chen, Q. Kang, M. Zhou, A. Abusorrah, and K. Sedraoui, "Dynamic embedding projection-gated convolutional neural networks for text classification," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 3, pp. 973–982, 2022.

[5] J. Bi, K. Xu, H. Yuan, J. Zhang, and M. Zhou, "Network attack prediction with hybrid temporal convolutional network and bidirectional GRU," *IEEE Internet Things J.*, vol. 11, no. 7, pp. 12 619–12 630, 2024.

[6] Z. Zhao, Z. Yang, C. Li, Q. Zeng, W. Guan, and M. Zhou, "Dual feature interaction-based graph convolutional network," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 9019–9030, 2023.

[7] M. Aljebreen, H. A. Mengash, M. Alamgeer, S. S. Alotaibi, A. S. Salama, and M. A. Hamza, "Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images," *IEEE Access*, vol. 12, pp. 11 147–11 156, 2024.

[8] K. Tong and Y. Wu, "Small object detection using deep feature learning and feature fusion network," *Eng. Appl. Artif. Intell.*, vol. 132, p. 107931, 2024.

[9] F. Zhu, J. Li, B. Zhu, H. Li, and G. Liu, "UAV remote sensing image stitching via improved VGG16 siamese feature extraction network," *Expert Syst. Appl.*, vol. 229, no. Part A, p. 120525, 2023.

[10] M. M. Hossen, M. E. Majid, S. B. A. Kashem, A. Khandakar, M. Nashbat, A. Ashraf, M. Hasan-Zia, A. K. A. Kunju, S. Kabir, and M. E. H. Chowdhury, "A reliable and robust deep learning model for effective recyclable waste classification," *IEEE Access*, vol. 12, pp. 13 809–13 821, 2024.

[11] K. Belsare, M. Singh, A. Gandam, P. K. Malik, R. Agarwal, and A. Gehlot, "An integrated approach of iot and WSN using wavelet transform and machine learning for the solid waste image classification in smart cities," *Trans. Emerg. Telecommun. Technol.*, vol. 35, no. 4, 2024.

[12] M. Molina, R. P. Ribeiro, B. Veloso, and J. Gama, "Super-resolution analysis for landfill waste classification," in *Advances in Intelligent Data Analysis XXII - 22nd International Symposium on Intelligent Data Analysis, IDA 2024, Stockholm, Sweden, April 24-26, 2024, Proceedings, Part I*, ser. Lecture Notes in Computer Science, I. Miliou, N. Piatkowski, and P. Papapetrou, Eds., vol. 14641. Springer, 2024, pp. 155–166.

[13] J. Hou, X. Xie, W. Wang, Q. Cai, Z. Deng, H. Yang, H. Huang, and Y. Wang, "Waste plastic bottles classification with deep learning model," *Int. J. Wirel. Mob. Comput.*, vol. 25, no. 3, pp. 296–302, 2023.

[14] S. Single, S. Iranmanesh, and R. Raad, "Realwaste: A novel real-life data set for landfill waste classification using deep learning," *Inf.*, vol. 14, no. 12, p. 633, 2023.

[15] D. Zheng, R. Wang, Y. Duan, P. C. Pang, and T. Tan, "Focus-rcnet: a lightweight recyclable waste classification algorithm based on focus and knowledge distillation," *Vis. Comput. Ind. Biomed. Art*, vol. 6, no. 1, p. 19, 2023.

[16] H. Jiang, Y. Lu, D. Zhang, Y. Shi, and J. Wang, "Deep learning-based fusion networks with high-order attention mechanism for 3d object detection in autonomous driving scenarios," *Appl. Soft Comput.*, vol. 152, p. 111253, 2024.

[17] J. Liang, H. Pan, Z. Xiang, J. Qin, Y. Qiu, L. Guo, T. Wang, W. Jiang, and B. Lei, "Echocardiographic segmentation based on semi-supervised deep learning with attention mechanism," *Multim. Tools Appl.*, vol. 83, no. 12, pp. 36 953–36 973, 2024.

[18] A. Kara, E. Pekel, E. Ozcetin, and G. B. Yildiz, "Genetic algorithm optimized a deep learning method with attention mechanism for soil moisture prediction," *Neural Comput. Appl.*, vol. 36, no. 4, pp. 1761–1772, 2024.

[19] D. Zhang, L. Lu, X. Li, J. Zhang, S. Zhang, and S. Yang, "Spatial downscaling of ESA CCI soil moisture data based on deep learning with an attention mechanism," *Remote. Sens.*, vol. 16, no. 8, p. 1394, 2024.

[20] P. Chen, H. Zhou, L. Ying, P. Liu, and B. Liu, "A novel deep learning network with deformable convolution and attention mechanisms for complex scenes ship detection in SAR images," *Remote. Sens.*, vol. 15, no. 10, p. 2589, 2023.

[21] W. Zhao, Z. Wu, D. Li, and Z. Yin, "Estimation of soil property content with vis-nir spectra by multitask deep learning based on attention mechanism and loss-weight balancing," *IEEE Geosci. Remote. Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[23] J. Chen, W. Wang, B. Liu, S. Cai, D. Towey, and S. Wang, "Hybrid semantics-based vulnerability detection incorporating a temporal convolutional network and self-attention mechanism," *Inf. Softw. Technol.*, vol. 171, p. 107453, 2024.

[24] Y. Xie, G. Liu, M. Zhou, L. Wei, H. Zhu, and R. Zhou, "A spatial-temporal gated network for credit card fraud detection," in *IEEE*

International Conference on Networking, Sensing and Control, ICNSC 2023, Marseille, France, October 25-27, 2023. IEEE, 2023, pp. 1–6.

[25] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 2018, pp. 4510–4520.