Efficient Rice Seed Classification Using Knowledge Distillation via ResNet18

Le Van Giap, Ho An Loc, Le Minh Dung
FPT University, Da Nang, Viet Nam

October 28 2024

Abstract

In recent years, rice variety classification has garnered attention from the scientific community due to its significant application potential in areas such as agricultural production, quality control, food security, and product traceability. Along with technological advancements, advanced Convolutional Neural Network (CNN) architectures have achieved substantial improvements in recognizing and classifying rice varieties. However, the increasing complexity of these models has led to larger sizes and slower prediction times, making it challenging to deploy on resource-constrained devices. The current research trend focuses on developing lightweight deep learning models for rice variety classification, aiming to maintain high accuracy while significantly reducing model size and computational requirements.

This study proposes a lightweight architecture based on the ResNet model, named RiceSeed-18 Lite, to enhance the efficiency of rice variety classification. RiceSeed-18 Lite allows users to select lighter network branches during the prediction process, adapting to the specific resource constraints of each scenario. Experimental results demonstrate that the proposed architecture outperforms the traditional ResNet18 model in terms of effectiveness.

I. Introduction

Rice is a staple food for more than half of the world's population, and the quality and purity of rice seeds are crucial for ensuring high yields and maintaining the genetic integrity of rice varieties. Traditionally, the classification of rice seeds has been performed manually by experts based on visual inspection. However, this method is labor-intensive, time-consuming, and subject to human error, which can lead to inconsistencies in seed quality assessment.

In recent years, deep learning techniques have been widely applied in agricultural applications, where precise classification and analysis of seeds are crucial for ensuring crop quality and yield. Among these, rice seed classification has emerged as an important area, as it helps in identifying seed varieties, detecting impurities, and ensuring quality control.

Traditional rice seed classification methods often require extensive manual labor and are prone to inconsistencies, prompting a shift towards automated, AI-driven solutions that can deliver higher accuracy and reliability. Convolutional Neural Networks (CNNs), particularly architectures like ResNet [3], have shown great promise in image classification due to their ability to learn rich, discriminative features. However, deploying such large models in practical, real-world scenarios, especially in rural and resource-constrained environments, presents significant challenges due to high computational and memory demands.

To address this issue, Knowledge Distillation (KD) [4] has emerged as a powerful technique, where a smaller "student" model learns from a larger, pretrained "teacher" model. This process enables the student model to retain the essential knowledge from the teacher, achieving comparable performance while being lighter and more efficient. By leveraging KD, models can be effectively scaled down for deployment on low-cost devices, which is especially valuable for agricultural applications where high-powered computing resources are not readily accessible.

In this work, we propose an enhanced rice seed classification approach based on a ResNet18 architecture integrated with Knowledge Distillation and attention mechanisms. Our method is designed to achieve high accuracy while remaining compact enough for deployment on devices with limited processing power. This enables farmers to utilize the model directly on affordable, portable devices without the need for high-end computing infrastructure, significantly reducing operational costs. By offering a fast, user-friendly, and cost-effective solution, our approach aims to make advanced AI technology accessible to the agricultural community, promoting efficient and accurate seed classification in real-world conditions.

II. Data pre-processing

a) Rice Seed Samples

Six commonly cultivated rice seed varieties in Northern Vietnam—BC-15, Huong Thom 1, Nep-87, Q-5, Thien uu-8, and Xi-23—were examined in [5]. These rice seeds were sourced from a production company where the varieties were cultivated and harvested under specific conditions to meet standard rice seed production requirements. The sampling was conducted in the Thai Binh and Ha Noi regions in Northern Vietnam.

b) Image Acquisition

We used a Nikon D300S CMOS color camera with a resolution of 640 x 480 pixels to capture images. The setup included a chamber with a white table serving as the background. Within a designated area of 10x16 cm^2, rice seeds were manually spread out. Each image taken by this system typically contained between 30 to 60 seeds. In total, we acquired 212 of these larger images. Our next step involves segmenting individual rice seed images from these acquired images.

c) Image segmentation

To separate individual rice seed images from the acquired images, we implemented image segmentation. Given that the image background is consistent across all experiments, we selected a threshold method for background subtraction. We found that the blue channel of the images had an intensity that effectively distinguished between the background and the rice seeds. Consequently, we used a threshold method based on the intensity values in the blue channel. Specifically, in the blue channel, the intensity of rice seed pixels is always 90 or less, while the intensity of the background is always greater than 90. During the image segmentation process, pixels with a blue value greater than 90 were assigned a value of 0, and pixels with a blue value of 90 or less were assigned a value of 255. After generating the threshold image, we cropped the rice seed images based on the object contours (Fig. 1), ensuring each image contains only one rice seed within a minimal bounding box. Hereafter, any reference to rice seed images pertains to this processed set of images.

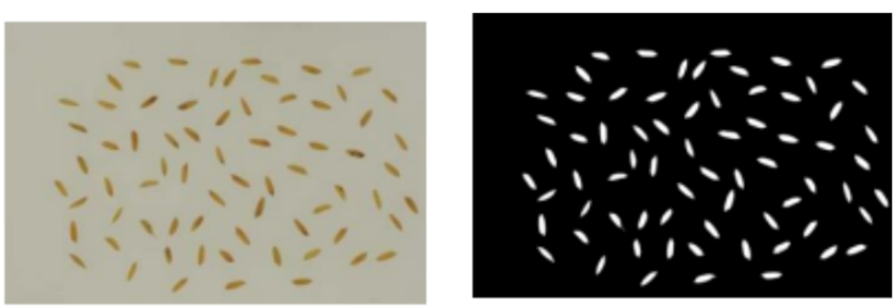

Figure 1: A sample of accquired image and its thresholded image.

III. MATERIALS AND METHODS

a) Overall Architecture

The proposed ResNet18-Lite model is divided into two components: the Teacher Branch at the top and the Student Branches, inspired by the work referenced in [2] below (Fig. 3. In the Teacher Branch, we utilize the original ResNet18 network. ResNet is a deep convolutional neural network architecture known for its effective use of skip connections, which help information and gradients flow directly through the layers, reducing gradient vanishing issues. This design has led to ResNet's success across numerous computer vision tasks, establishing it as one of the most reliable and widely used CNN architectures. The structure of ResNet typically includes:

- An initial 7x7 convolution layer.
- A max pooling layer
- Four residual blocks, each composed of convolutional layers and residual connections.
- A global average pooling layer.
- A final fully connected layer.

b) Distillation Block

A Distillation Block, customized for rice seed classification, combines Depth-wise Separable Convolution layers [1] with skip connections (Fig. 2). This core component is designed to distill critical information from the Teacher Branch while preserving essential data features. Each Depth-wise Separable Convolution convolution layer within the block has filters of sizes [64, 128, 256], with a stride of 2. By stacking these distillation blocks and doubling the number of filters after each block, the model enhances its learning capacity to capture more complex features relevant to distinguishing rice seed varieties.

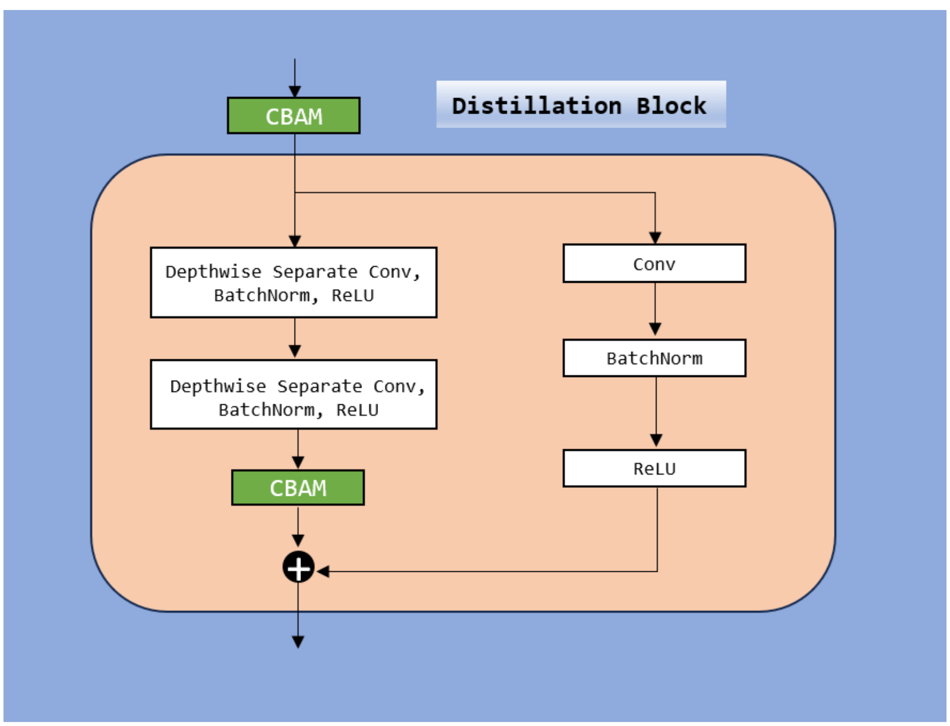

Figure 2: Structure of a distillation block consisting of Depth-wise Separable Convolution layers and skip connections

Using Depth-wise Separable Convolution instead of standard convolutions significantly reduces the number of parameters, making the model lighter and more suitable for deployment on devices with limited resources. Additionally, integrated skip connections facilitate efficient back-propagation through network layers, helping prevent information loss and enabling more effective feature learning.

To further improve feature selection, we apply the Convolutional Block Attention Module (CBAM) [6]. CBAM, consisting of Channel Attention and Spatial Attention components in sequence, helps the model focus on important regions of the image. In rice seed classification, certain visual features are more significant than others for distinguishing varieties, and CBAM allows the model to dynamically adjust its focus to these critical features.

c) Training method

Given N samples $X = \{x_i\}_{i=1}^N$ from 2 classes, with the corresponding label set $Y = \{y_i\}_{i=1}^N$. The proposed ResNet18-Lite architecture includes K classifiers (where $K = 3$), denoted as $C = \{c_1, c_2, c_3\}$. Classifiers c_1, c_2, and c_3 are positioned in the Student branches of the network. For each vector z^k output after the fully connected layers of classifier c_k, the probability distribution based on the sigmoid function of the vector z^k after softening with temperature T is calculated as follows:

$$q_i^k = \frac{1}{1 + \exp(-z_i^k/T)}$$

The training process is divided into two branches:

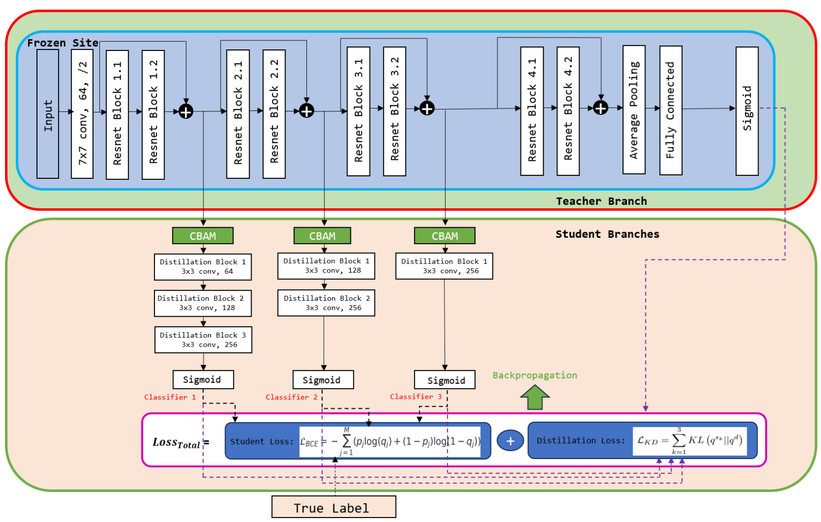

Figure 3: Overall architecture of the proposed model

- **The Teacher Branch** uses the Binary Cross-Entropy (BCE) loss function to adjust the model weights. This loss function encourages the model to assign high probabilities to the correct class and low probabilities to the incorrect class. It is defined as follows:

$$L_{BCE} = -\sum_{j=1}^{2} \left(p_j \log(q_j) + (1 - p_j) \log(1 - q_j)\right)$$

  where p_j is the true label and q_j is the sigmoid probability for class j.

- **The Student Branch** learns simultaneously from the input data and the Teacher Branch. The general loss function for the Student Branch is defined as a combination of Binary Cross-Entropy Loss (L_{BCE}), Distillation Loss (L_{KD}). The Distillation Loss (L_{KD}) [10] is used to align the outputs of the student branches with those of the teacher branch. This is achieved by minimizing the difference between the sigmoid probabilities of the teacher branch q_d and those of the student branches q_s. This loss is formulated as:

$$L_{KD} = \sum_{k=1}^{3} KL(q_{s_k}||q_d)$$

where q_s is the sigmoid probability of the student branch classifiers c_1, c_2, and c_3, and q_d is the sigmoid probability of the teacher branch classifier.

- The importance of each loss component is controlled by two balancing parameters, α **and** λ. The total loss function is defined as:

$$L_{Total} = (1 - \alpha) \cdot L_{BCE} + \alpha \cdot L_{KD}$$

where α governs the balance between the Binary Cross-Entropy loss and the Knowledge Distillation loss. In our experiments, we set $\alpha = 0.1$ and the temperature $T = 3$.

IV. Experiments

a) Experimental Results and Evaluation

	ResNet18	**C1**	**C2**	**C3**
BC-15	97.69%	96.87%	98.09%	97.96%
Huong Thom	98.07%	99.15%	99.39%	99.15%
Nep-87	97.74%	96.88%	97.05%	96.53%
Q-5	94.03%	94.52%	94.19%	95.19%
Thien Uu-8	97.51%	97.76%	97.76%	98.75%
Xi-23	96.02%	97.95%	97.46%	97.10%
Average	96.84%	97.19%	97.32%	97.45%

Table 1: Accuracy of Different Models on Rice Varieties

Model	**Trainable parameters**	**Total parameters**
ResNet18	11,181,825	11,191,425
C1	1,070,733	1,231,885
C2	1,020,202	1,708,458
C3	822,983	3,614,023

Table 2: Trainable and Total Parameters for Different Models

The models C2 and C3 achieved the highest average accuracy on rice variety classification, with C3 reaching 97.45% and C2 at 97.32%. This indicates their effectiveness in distinguishing rice varieties, surpassing ResNet18's average accuracy of 96.84% . Despite ResNet18's high computational demands due to its over 11 million parameters, the smaller architectures(C1, C2, and C3) demonstrate competitive or even superior performance. Notably, C1, with a parameter count of only 1.07 million, reached an accuracy of 97.19%, showing that it remains a viable option for applications where reduced parameters are essential. C3 stands out as it combines the fewest parameters with the highest accuracy, suggesting that its structure is optimally designed for rice classification, effectively balancing performance with reduced computational overhead.

Table 3 presents the effectiveness of the models when using combined loss, compared to the results in Table 1, which uses standard loss functions. Under the combined loss approach, models C2 and C3 achieved slightly higher accuracy than when using standard loss, with Combine C3 reaching the highest average accuracy of 97.78% and Combine C2 following closely at 97.42%. These results suggest that applying combined loss can enhance model performance across various architectures, particularly smaller ones like C2 and C3, compared to standard loss. The accuracy of Combine C1, at 93.52%, provides an additional perspective on the model's performance under combined loss conditions relative to its standard loss result in Table 1.

	Combine C1 accuracy	**Combine C2 accuracy**	**Combine C3 accuracy**
BC-15	96.16%	98.58%	98.07%
Huong Thom	97.24%	98.88%	99.68%
Nep-87	95.58%	98.70%	98.23%
Q-5	93.36%	95.03%	95.23%
Thien Uu	86.62%	97.28%	98.97%
Xi-23	92.16%	96.04%	96.50%
Average	93.52%	97.42%	97.78%

Table 3: Accuracy of Different Models on Rice Varieties (Combine loss in C1, C2, C3)

V. Conclusion

This study introduced **ResNet18-Lite**, a lightweight and efficient model that leverages the principles of knowledge distillation while integrating Distillation Blocks with Depth-wise Separable Convolution and the Convolutional Block Attention Module (CBAM) attention mechanisms. By doing so, ResNet18-Lite significantly enhances feature extraction and knowledge transfer within the model, achieving notable improvements over its predecessor, the recent variant of ResNet18.

Despite having a reduced number of parameters characteristic of lightweight models, ResNet18-Lite offers remarkable performance and efficiency. This flexibility allows users to select the most suitable network branches according to their specific resource constraints. Notably, the shallowest branch of our architecture, despite containing the fewest parameters, maintains a high level of performance, demonstrating that even compact models can deliver effective results in various applications. The advancements presented in ResNet18-Lite highlight the potential for lightweight architectures to perform competitively while being mindful of computational resource limitations.

References

[1] François Chollet. Xception: Deep learning with depthwise separable convolutions, 2017.

[2] Huy-Hoang Dinh, Hong-Quan Do, Trung-Tung Doan, Cuong Le, Ngo Xuan Bach, Tu Minh Phuong, and Viet-Vu Vu. Fgw-fer: Lightweight facial expression recognition with attention. *KSII Transactions on Internet and Information Systems (TIIS)*, 17(9):2505–2528, 2023.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[4] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015.

[5] Thi-Thu-Hong Phan, Tran Thi Thanh Hai, Thi-Lan Le, Vo Ta Hoang, Hai Vu, and Thuy Thi Nguyen. Comparative study on vision based rice seed varieties identification. *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, pages 377–382, 2015.

[6] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module, 2018.