

Frequency-Regularized Partial Heat Routing for Efficient Pollen Recognition

Hongyu Song

Inner Mongolia University of Technology

Bao Shi

kshibao@imut.edu.cn

Inner Mongolia University of Technology

Jingping Yang

Inner Mongolia Baogang Hospital

XinRu Zhang

Inner Mongolia University of Technology

Yuan Xue

Inner Mongolia University of Technology

Mengfei Wang

Inner Mongolia University of Technology

Article

Keywords: Pollen recognition, Computational microscopy, Efficient vision backbones, Heat Conduction Operator, Frequency-domain regularization, Accuracy–efficiency trade-of

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10724323/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Redundancy-Aware Frequency and Routing Optimization for Efficient Fine-Grained Pollen Recognition

Hongyu Song¹, Bao Shi^{2*}, Jingping Yang³, XinRu Zhang¹,
Yuan Xue¹, Mengfei Wang¹

¹College of Intelligent Science and Technology, Inner Mongolia
University of Technology, Hohhot, Inner Mongolia, China.

²School of Information Engineering, Inner Mongolia University of
Technology, Huhhot, Inner Mongolia, China.

³Inner Mongolia Baogang Hospital, Baotou, Inner Mongolia, China.

*Corresponding author(s). E-mail(s): kshibao@imut.edu.cn;
Contributing authors: 20241800119@imut.edu.cn;

Abstract

Airborne pollen recognition is important for automated allergy warning and environmental monitoring, but remains challenging because inter-class morphological differences are subtle while intra-class appearances vary considerably. To improve the accuracy–efficiency trade-off of pollen recognition, we investigate redundancy within the Heat Conduction Operator (HCO) of the heat-based vision backbone vHeat. Channel-path and DCT-domain frequency-response analyses reveal exploitable optimization space in HCO. Based on these observations, we propose a redundancy-aware frequency and routing optimization framework consisting of Partial Heat Routing (PHR), Frequency-domain Regularization (FDR), and a stage-wise deployment strategy. PHR routes only a subset of channels through the complete Heat2D branch and processes the remaining channels with a lightweight local branch, reducing full-channel heat-conduction computation. FDR introduces a controlled frequency-retention constraint in the DCT domain, while the stage-wise strategy coordinates PHR and FDR across network depths. On Pollen149, the final configuration reduces parameters from 30.20M to 25.69M and FLOPs from 4.94G to 3.97G with only a 0.19 percentage-point decrease in Top-1 accuracy. Throughput increases from 924.15 to 1015.27 images/s and

latency decreases from 1.0821 to 0.9850 ms. Additional public-dataset experiments further indicate competitive recognition performance under different data conditions.

Keywords: Pollen recognition, Computational microscopy, Efficient vision backbones, Heat Conduction Operator, Frequency-domain regularization, Accuracy–efficiency trade-off

1 Introduction

Pollen allergy has become an increasingly important public health issue worldwide. Airborne allergenic pollen is closely associated with respiratory allergic diseases, and its production, release, distribution, monitoring, forecasting, and health impacts have been widely investigated [1]. Recent studies further indicate that climate change has intensified pollen seasons, making reliable and timely pollen monitoring increasingly important [2]. Accurate monitoring of airborne pollen concentrations is essential for public health warning, allergy prevention, and medical resource allocation, with precise pollen identification being a critical step in this process. Currently, conventional pollen monitoring and identification still largely rely on manual microscopic analysis by trained professionals, which is labor-intensive and difficult to scale for rapid reporting [3]. Pollen grains also exhibit high inter-class similarity and substantial intra-class variation. Different species may have highly similar appearances, whereas samples from the same species may vary in size, shape, surface texture, and aperture structure. Pollen recognition is therefore a challenging fine-grained microscopic image recognition task that requires models to distinguish subtle morphological differences.

Deep learning methods have been widely applied to pollen image classification [4–6]. On several pollen datasets, convolutional neural networks (CNNs) have achieved better recognition performance than traditional methods based on hand-crafted features. The locality of convolution, however, limits their ability to capture long-range dependencies, whereas pollen recognition often requires the joint representation of local micro-textures and global morphological structures. Vision Transformers alleviate this limitation by establishing global information interactions through self-attention [7–9]. Standard self-attention has a computational complexity that grows quadratically with the number of image tokens, leading to considerable cost when processing high-resolution microscopic images. Efficient pollen recognition therefore requires visual backbones that balance global morphology modeling, local microscopic detail representation, and practical inference efficiency.

In this context, heat-conduction-inspired visual modeling provides an alternative approach to efficient global token interaction. vHeat introduces the Heat Conduction Operator (HCO), which models spatial interactions through a heat-diffusion-like process in the frequency domain [10]. Unlike self-attention, HCO does not explicitly compute pairwise token interactions. Instead, it combines the discrete cosine transform (DCT), frequency-domain heat-diffusion modulation, and inverse DCT reconstruction to achieve global feature interaction with favorable computational efficiency. This

mechanism is well aligned with the requirements of pollen recognition for global morphological structures, local microscopic details, and practical inference efficiency. We therefore adopt vHeat as the base backbone and investigate how its computational cost can be further reduced while largely preserving its global representation capability.

Although vHeat achieves efficient global feature interaction through HCO, its internal computation may still contain exploitable redundancy. A complete HCO applies the Heat2D transformation to all channels and modulates the resulting frequency responses in the DCT domain before reconstruction. Different channels and frequency responses do not necessarily contribute equally to the final recognition result. We therefore analyze HCO from two perspectives: channel computation paths and DCT-domain frequency responses. Channel-response statistics and masking experiments indicate that different channels rely on the complete Heat2D transformation to different degrees. Frequency-response analysis further shows that some responses in the middle-to-late stages can be moderately constrained with limited performance degradation. These observations suggest optimizable computation in both the channel path and frequency-response dimensions, providing empirical motivation for the subsequent method design. They do not imply that weak-response channels or particular frequency components should be directly removed; instead, the corresponding redundancy should be exploited in a controlled manner.

Based on these observations, we propose an HCO-oriented redundancy-aware frequency and routing optimization framework to improve the accuracy–efficiency trade-off of vHeat for fine-grained pollen recognition. The framework consists of Partial Heat Routing (PHR), Frequency-domain Regularization (FDR), and a stage-wise deployment strategy. PHR restructures the channel computation path of HCO by sending only a subset of channels through the complete Heat2D branch. The remaining channels are processed by a lightweight local branch, reducing full-channel heat-conduction computation while preserving the output channel dimensionality. FDR introduces a controlled frequency-retention constraint in the DCT domain of HCO to regulate the use of different frequency responses. Different stages of the hierarchical vHeat backbone differ in spatial resolution, channel dimension, computational contribution, and semantic level. We therefore design a stage-wise deployment strategy that assigns different PHR routing ratios and FDR frequency-retention ratios across network depths. The redundancy analyses guide the design of candidate configurations, while the final configuration is determined through systematic ablation experiments according to the accuracy–efficiency trade-off.

We use Pollen149 as the primary evaluation dataset and further validate the proposed framework on public pollen datasets. On Pollen149, the final stage-wise PHR+FDR configuration reduces the number of parameters from 30.20M to 25.69M and FLOPs from 4.94G to 3.97G, with only a 0.19 percentage-point decrease in Top-1 accuracy. Under the unified inference protocol, throughput increases from 924.15 images/s to 1015.27 images/s, while latency decreases from 1.0821 ms to 0.9850 ms. Additional experiments on public pollen datasets show that the proposed framework maintains competitive recognition performance under different data distributions and imaging conditions. These results indicate that the proposed optimization improves

the accuracy–efficiency trade-off of vHeat while largely preserving the recognition capability of the original backbone.

The main contributions of this paper are summarized as follows:

- **We design Partial Heat Routing to reduce redundant full-channel heat-conduction computation.** PHR sends only a subset of channels through the complete HCO branch and processes the remaining channels using a lightweight local branch. It preserves the output channel dimensionality while reducing full-channel Heat2D computation, providing a structured approach to optimizing channel computation paths within HCO.
- **We introduce Frequency-domain Regularization and design a stage-wise deployment strategy to coordinate PHR and FDR.** FDR imposes a controlled frequency-retention constraint in the DCT domain of HCO to regulate the use of different frequency responses. The stage-wise strategy further assigns different PHR routing ratios and FDR frequency-retention ratios at different network depths, coordinating structural computation optimization with frequency-domain representation constraints.
- **We validate the accuracy–efficiency trade-off of the proposed framework on Pollen149 and public pollen datasets.** The final stage-wise PHR+FDR configuration reduces model parameters and FLOPs and improves practical inference efficiency while maintaining recognition performance close to that of the original vHeat. Experiments on public pollen datasets further show competitive performance under different data distributions and imaging conditions.

2 Related Work

2.1 Automated Pollen Recognition and Microscopic Image Analysis

Early automated pollen recognition methods mainly relied on hand-crafted morphological, color, texture, and statistical features, such as local binary patterns, gray-level co-occurrence matrices, histograms of oriented gradients, Fourier descriptors, and Haralick features. These features were typically combined with conventional classifiers, including support vector machines, linear discriminant analysis, k -nearest neighbors, random forests, and multilayer perceptrons [4]. Although such methods provide a certain degree of morphological interpretability, their performance depends heavily on manual feature design, limiting their ability to represent subtle inter-class differences and complex intra-class variations.

With the development of deep learning, convolutional neural networks, residual networks, lightweight architectures, and pretrained visual models have been increasingly applied to microscopic pollen image classification and extensively evaluated on datasets such as POLEN23E, POLLEN13K, CPD, Pollen73S, and regional pollen collections [5, 6, 11–13]. These studies have shown that deep visual models can effectively learn discriminative features related to overall morphology, boundary structures, and surface textures, while supporting automated or online pollen recognition. For Z-stack multi-focal microscopic images, three-dimensional convolutional networks have

also been employed to integrate complementary information across different focal planes [14]. More recently, attention mechanisms and Transformer architectures have been introduced to enhance the joint modeling of global morphology, local textures, aperture-related structures, and long-range spatial relationships [8, 9, 15].

Beyond the classification of cropped single-grain images, task-specific networks, explainable recognition methods, and detection–classification pipelines have also been developed for complex microscopic-image analysis. These approaches employ dedicated classification architectures or visualization techniques to identify discriminative pollen regions, or first use YOLO-based detectors to localize individual pollen grains in images with complex backgrounds, overlapping objects, or multiple targets, followed by classification using CNN, SVM, or Transformer-based models [16–18]. Overall, pollen image analysis has evolved from conventional classification based on hand-crafted features toward deep representation learning, attention-enhanced modeling, and automated frameworks integrating detection and classification.

2.2 Efficient Visual Modeling with Partial Feature Processing

Efficient visual models commonly reduce parameter counts and computational costs through lightweight operators and compact network architectures. The MobileNet family adopts depthwise separable convolutions and inverted residual structures, while the ShuffleNet family improves computational efficiency through group convolution and channel shuffling. EfficientNet further balances accuracy and efficiency by jointly scaling network depth, width, and input resolution [19–25]. In addition, GhostNet generates additional feature maps through inexpensive transformations, whereas hybrid convolution–Transformer architectures combine local operations, lightweight attention, and hardware-friendly designs to improve inference efficiency while maintaining representational capability [26–30].

Partial feature processing further reduces redundant computation by partitioning feature flows rather than applying the same full-cost transformation to all features. For example, CSPNet partitions cross-stage feature flows to alleviate duplicated gradient information, whereas GhostNet generates additional representations from a small set of intrinsic features using inexpensive operations [26, 31]. Related approaches include dynamic inference, which adaptively allocates computational resources through early exiting or layer- and channel-level selection, and model-compression techniques, such as pruning, quantization, knowledge distillation, and low-rank decomposition, which reduce the storage or computational cost of existing networks [32–41]. These studies demonstrate that assigning different computational paths to different features can reduce repeated processing and improve the accuracy–efficiency trade-off. However, existing designs mainly focus on convolutional feature generation, cross-stage connections, or general model compression, while comparatively limited attention has been paid to channel-wise computation paths within frequency-domain global operators.

2.3 Frequency-Domain Visual Modeling and Heat-Conduction-Based Networks

Frequency-domain visual modeling transforms spatial features into the frequency domain through the Fourier transform or discrete cosine transform, enabling global information interaction or selective frequency processing. GFNet replaces self-attention with two-dimensional Fourier transforms, learnable global filters, and inverse transforms to model long-range dependencies with relatively low computational complexity. DCFormer directly learns visual semantics from DCT-domain representations and reduces computational costs by compressing selected frequency components [42, 43]. In addition, frequency-aware dynamic networks allocate computational paths of different complexities to image regions according to their DCT-frequency information, further demonstrating the potential of frequency decomposition for balancing representation quality and computational cost [44].

Building upon frequency-domain visual modeling, vHeat introduces the Heat Conduction Operator (HCO), which models image patches as heat sources and performs adaptive heat diffusion in the DCT domain to achieve global feature interaction, thereby constructing heat-conduction-based visual backbones [10]. RS-vHeat subsequently extends HCO to remote-sensing foundation models and combines it with frequency-domain hierarchical masking and multi-domain reconstruction for remote-sensing representation learning [45]. These studies demonstrate that frequency-domain transformations and heat-conduction mechanisms provide an alternative approach to efficient global visual modeling beyond convolution and self-attention.

3 Method

3.1 Framework Overview and Design Rationale

The optimization target of this work is the Heat Conduction Operator (HCO) in vHeat [10]. The proposed framework is HCO-oriented and does not modify the overall classification pipeline of vHeat. Instead, it focuses on reducing exploitable redundancy within HCO while preserving its heat-diffusion-based representation capability. Given an input feature U^0 , the basic formulation of HCO can be written as

$$U^t = \text{IDCT}_{2D} \left(\text{DCT}_{2D}(U^0) \cdot e^{-k(\omega_x^2 + \omega_y^2)t} \right), \quad (1)$$

where U^0 and U^t denote the visual features before and after heat conduction, respectively. ω_x and ω_y are the frequency coordinates, k is the heat diffusion coefficient, and t denotes the conduction time. HCO first transforms the input feature into the DCT domain, modulates the frequency responses using a heat-diffusion term, and then reconstructs the spatial feature through IDCT. This process is applied to all channels in the complete Heat2D branch. It therefore provides two natural perspectives for redundancy-aware optimization: the channel computation path and the DCT-domain frequency-response distribution.

From the channel-path perspective, the original HCO applies the same DCT, frequency-domain modulation, and IDCT operations to all channels. This full-channel

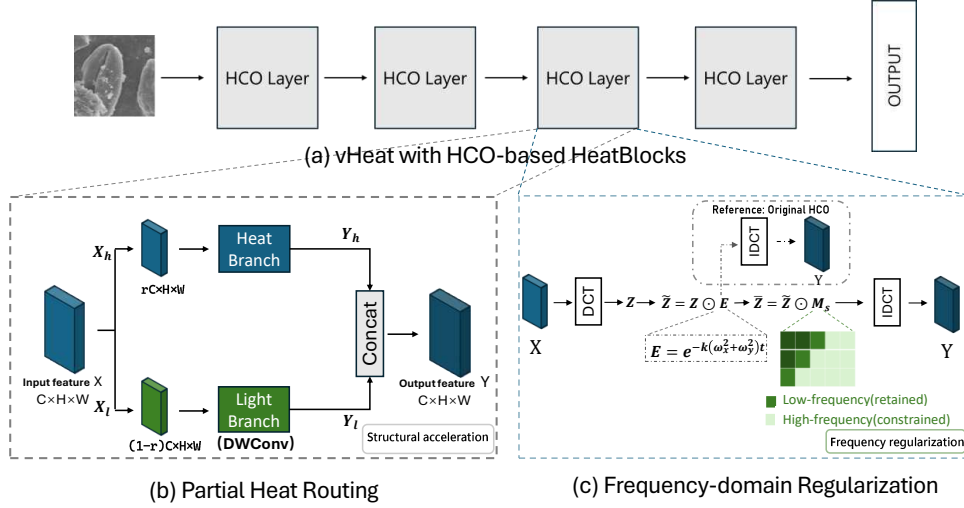


Fig. 1 Overall framework of the proposed HCO-oriented redundancy-aware frequency and routing optimization framework. (a) The vHeat backbone uses stacked HCO-based HeatBlocks for global token mixing. (b) Partial Heat Routing (PHR) splits the input feature into a Heat branch with rC channels and a lightweight local branch with $(1-r)C$ channels, and concatenates their outputs to preserve the original channel dimension. (c) Frequency-domain Regularization (FDR) inserts a stage-wise frequency retention mask M_s after the heat-diffusion modulation in the DCT domain. The reference path shows that the original HCO directly reconstructs the spatial feature from $\tilde{Z}$ without FDR.

design preserves the modeling capacity of heat diffusion but also concentrates substantial computation in the complete Heat2D branch. Different channels may not depend on the complete Heat2D transformation to the same extent. Some channels may respond weakly to heat diffusion or retain sufficient local representation ability through a lightweight transformation. These observations motivate Partial Heat Routing (PHR), which reallocates part of the channels from the complete HCO branch to a lightweight local branch.

From the frequency-response perspective, HCO modulates different frequency components in the DCT domain. Empirical analysis indicates that some frequency responses in the middle-to-late stages can be moderately constrained with limited performance degradation, whereas overly strong suppression may harm recognition performance. We therefore do not treat these frequency responses as direct targets for structural compression. Instead, they provide an entry point for representation constraint. Based on this idea, we design Frequency-domain Regularization (FDR), which introduces a controlled frequency-retention constraint inside HCO to regulate the use of different frequency responses.

Based on the above analysis, the proposed HCO-oriented redundancy-aware optimization framework consists of two components and one deployment strategy. PHR

reduces redundant full-channel Heat2D computation by reallocating channel computation paths while preserving the output channel dimensionality. FDR introduces a controlled frequency-domain constraint to regulate the frequency-response distribution inside HCO. Since vHeat adopts a hierarchical architecture, its stages differ in spatial resolution, channel dimension, computational contribution, and semantic level. We therefore further introduce a stage-wise deployment strategy to coordinate PHR and FDR across different network depths.

The redundancy analyses provide empirical guidance for the design of candidate configurations. The final stage-wise configuration is determined through ablation experiments according to the accuracy–efficiency trade-off. The overall framework is illustrated in Fig. 1. The following subsections describe PHR, FDR, and the stage-wise deployment strategy in detail.

3.2 Partial Heat Routing for Channel-Path Optimization

Within the proposed HCO-oriented redundancy-aware optimization framework, Partial Heat Routing (PHR) is designed to reduce redundant full-channel Heat2D computation by reallocating channel computation paths. The core idea is to split the input feature along the channel dimension into two branches. A subset of channels is processed by the complete HCO branch to preserve heat-diffusion-based global interaction. The remaining channels are processed by a lightweight local branch at lower computational cost. As illustrated in Fig. 1(b), the outputs of the two branches retain the original channel dimensionality. In this way, PHR reduces the computational burden of the complete Heat2D branch without changing the interface of the original HeatBlock.

Given an input feature $X \in \mathbb{R}^{B \times C \times H \times W}$, where B , C , H , and W denote the batch size, number of channels, height, and width, respectively, PHR uses a heat ratio r to control the proportion of channels entering the complete HCO branch. The input feature is split along the channel dimension into a Heat sub-feature and a lightweight sub-feature:

$$X = [X_h, X_l], \quad X_h \in \mathbb{R}^{B \times C_h \times H \times W}, \quad X_l \in \mathbb{R}^{B \times C_l \times H \times W}, \quad (2)$$

where

$$C_h = \text{Round}(rC), \quad C_l = C - C_h. \quad (3)$$

Here, X_h is fed into the complete HCO branch, whereas X_l is routed to the lightweight branch. The heat ratio r controls the extent of partial channel routing. When $r = 1.0$, all channels enter the complete HCO branch and the structure recovers the original HeatBlock. When $r < 1.0$, only C_h channels undergo the complete heat-conduction transformation, while the remaining C_l channels are processed by the lightweight local branch.

For the Heat branch, we retain the original HCO structure of vHeat:

$$Y_h = \mathcal{H}(X_h), \quad (4)$$

where $\mathcal{H}(\cdot)$ denotes the Heat Conduction Operator composed of DCT, heat-diffusion modulation, and IDCT. Since this branch operates only on C_h channels, the computational cost of the complete HCO is reduced compared with the original full-channel design.

The remaining channels are processed by a lightweight local transformation:

$$Y_l = \mathcal{L}(X_l), \quad (5)$$

where $\mathcal{L}(\cdot)$ denotes the lightweight branch. In our implementation, this branch consists of a depthwise 3×3 convolution followed by a GELU activation:

$$\mathcal{L}(X_l) = \text{GELU}(\text{DWConv}_{3 \times 3}(X_l)). \quad (6)$$

This branch provides local feature transformation at substantially lower computational cost than the complete HCO branch.

The outputs of the two branches are concatenated along the channel dimension:

$$Y = \text{Concat}(Y_h, Y_l). \quad (7)$$

The resulting feature Y contains the same number of channels as the input X . PHR therefore preserves the original channel dimensionality while reallocating computation between the Heat and lightweight branches.

Inside the HeatBlock, the concatenated feature is further processed by normalization, residual connection, and the MLP branch. This process can be simplified as

$$\hat{X} = X + \text{DropPath}(\text{Norm}(Y)), \quad (8)$$

$$X^{\text{out}} = \hat{X} + \text{DropPath}(\text{Norm}(\text{MLP}(\hat{X}))). \quad (9)$$

PHR can therefore be embedded into the original HeatBlock without changing the hierarchical architecture or classification head of vHeat. Different routing ratios are further combined with the stage-wise deployment strategy and evaluated in the ablation experiments to determine a suitable accuracy-efficiency trade-off.

3.3 Frequency-domain Regularization

Within the proposed framework, Frequency-domain Regularization (FDR) is introduced to constrain the DCT-domain frequency responses inside HCO. Unlike PHR, which changes the channel computation path, FDR does not determine whether a channel enters the HCO branch. Instead, it directly acts on the frequency representation within HCO. As illustrated in Fig. 1(c), FDR inserts a frequency-retention mask after heat-diffusion modulation and before IDCT reconstruction, thereby applying a controlled constraint to part of the frequency responses.

Given an input feature $X \in \mathbb{R}^{B \times C \times H \times W}$ fed into HCO, a two-dimensional DCT is first applied to obtain the frequency-domain representation:

$$Z = \text{DCT}_{2D}(X), \quad (10)$$

where $Z \in \mathbb{R}^{B \times H_f \times W_f \times C}$ denotes the frequency responses in the DCT domain, and H_f and W_f denote the frequency-domain height and width, respectively. HCO then modulates these responses using the adaptive heat-diffusion coefficients generated by Frequency Value Embeddings:

$$\tilde{Z} = Z \odot E, \quad (11)$$

where E denotes the frequency-domain modulation weights formed by the heat-diffusion term, and $\odot$ denotes element-wise multiplication. In the original vHeat, $\tilde{Z}$ is directly fed into IDCT for spatial-domain reconstruction. FDR introduces an additional frequency-retention mask before this reconstruction step:

$$\bar{Z} = \tilde{Z} \odot M_s, \quad (12)$$

where $M_s \in \mathbb{R}^{H_f \times W_f \times 1}$ denotes the frequency-domain regularization mask for the s -th stage.

Since low-frequency responses are concentrated in the upper-left region of the DCT spectrum, M_s is constructed using a low-frequency retention scheme. Given a stage-wise frequency-range retention ratio ρ_s , the mask is defined as

$$M_s(i, j) = \begin{cases} 1, & i < \text{Round}(\rho_s H_f), j < \text{Round}(\rho_s W_f), \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Here, $\rho_s \in (0, 1]$ controls the retained frequency-response range in the s -th stage. When $\rho_s = 1.0$, all frequency responses are preserved and HCO remains consistent with the original vHeat. When $\rho_s < 1.0$, the model mainly retains the low-frequency region while constraining part of the higher-frequency responses. The regularized frequency representation $\bar{Z}$ is then transformed back to the spatial domain through IDCT:

$$Y = \text{IDCT}_{2D}(\bar{Z}). \quad (14)$$

FDR does not reduce the number of channels entering the complete HCO branch. Instead, it regulates the frequency-response distribution within HCO. The preceding frequency analysis indicates that some middle-to-late-stage responses can be moderately constrained with limited performance degradation, whereas overly strong frequency suppression may impair recognition performance. FDR therefore uses an adjustable frequency-retention range to introduce a controlled representation constraint into the frequency-domain modeling process rather than directly removing structural components.

The dependence on frequency information also varies across network stages. Shallow features contain more edges, textures, and local details and are generally more sensitive to high-frequency information. Deeper features gradually become more related to structural and category-level representations. Based on this characteristic, FDR adopts different frequency-retention settings across stages. Conservative settings are used to preserve shallow details, while moderate frequency constraints are applied in the middle-to-late stages. The specific stage-wise configuration is described in the following subsection.

3.4 Stage-wise Deployment Strategy

PHR and FDR optimize different aspects of HCO. PHR adjusts the channel computation path to reduce full-channel Heat2D computation, whereas FDR regulates the DCT-domain frequency responses through a frequency-retention constraint. The hierarchical stages of vHeat differ in spatial resolution, channel dimension, computational contribution, and semantic level. A uniform routing or frequency-retention setting may therefore be suboptimal across all stages. To account for these differences, we introduce a stage-wise deployment strategy that coordinates PHR and FDR across network depths.

For the s -th stage, PHR is controlled by the channel routing ratio r_s , which denotes the proportion of channels entering the complete Heat2D branch. A smaller r_s reduces the number of channels processed by the complete HCO branch and lowers the structural computational cost. An overly small value, however, may weaken heat-diffusion-based global modeling. FDR is controlled by the frequency-range retention ratio ρ_s , which determines the retained range of DCT-domain frequency responses. A smaller ρ_s imposes a stronger frequency-domain constraint, whereas excessive suppression may remove useful discriminative information.

The two ratios serve different functional objectives and are therefore configured separately. For FDR, conservative frequency-retention settings are used in the early stages to preserve edges, textures, and local details. Stronger constraints are applied in the middle-to-late stages, where the frequency-response analysis indicates greater room for controlled regularization. PHR mainly targets structural computational efficiency. Its routing ratios are selected by considering the computational contribution of each stage, the channel-path analysis, and the corresponding changes in recognition performance.

The redundancy analyses guide the design of candidate stage-wise configurations rather than directly determining the final ratios. The specific combinations of r_s and ρ_s are evaluated through ablation experiments, and the final configuration is selected according to the accuracy–efficiency trade-off. The detailed stage-wise settings and corresponding experimental results are reported in the experimental section.

4 Experiments

4.1 Experimental Setup

We use Pollen149 as the primary experimental dataset to systematically evaluate the influence of PHR and FDR on the accuracy–efficiency trade-off of the vHeat backbone. To further examine the applicability of the proposed method under different data sources, we conduct additional validation on two public pollen datasets: the New Zealand Pollen Dataset [46] and CPD-1 [47]. The main experiments focus on recognition accuracy, model complexity, and practical inference efficiency. The public dataset experiments are used to further examine the behavior of the proposed method under different pollen data distributions and imaging conditions.

Pollen149 is an in-house pollen microscopy dataset previously established by our research group using scanning electron microscopy (SEM). The pollen materials used

to construct the dataset were provided by the Institute of Biomedical Research, Inner Mongolia University, Hohhot, Inner Mongolia, China, together with the corresponding taxonomic information. The species labels in Pollen149 were assigned according to this existing taxonomic information, and no additional independent taxonomic identification was performed by our research group. The samples were subsequently prepared and imaged using SEM at Inner Mongolia University of Technology between 2018 and 2024. The resulting images were organized into the Pollen149 dataset, which contains 29,018 images with a unified resolution of 224×224 , covering 51 families and 149 species.

No wild plant or pollen material was collected by our research group during the construction of Pollen149; therefore, no field-collection permit was required for the activities conducted by our group. No voucher specimens were deposited in a public herbarium by our research group during dataset construction. The present study uses only the previously established SEM image dataset and does not involve any new collection of plant or pollen materials. All sample handling and SEM imaging procedures conducted by our research group complied with the applicable institutional, national, and international guidelines and regulations.

Differences among pollen categories are often reflected in subtle variations in local texture, aperture structures, grooves, and overall morphology. Pollen149 therefore exhibits substantial intra-class variation and high inter-class similarity, making it suitable for evaluating fine-grained pollen recognition methods. We split the dataset into training, validation, and test sets at a ratio of 4:1:1. The same split files are used for all compared models and ablation variants.

The New Zealand Pollen Dataset contains 46 categories and 19,667 darkfield microscopy images, with substantial class imbalance. CPD-1 contains 4,034 segmented pollen grain images from 20 pollen species collected in Crete, Greece. In our experiments, both public datasets are split into training, validation, and test sets at a ratio of 7:1:2 using random seed 42. Stratified splitting is used when the class sample sizes allow it. These datasets differ from Pollen149 in imaging modality, category composition, sample scale, and data distribution, and are therefore used to further evaluate the applicability of the proposed method under different pollen data conditions.

All datasets are loaded from predefined CSV split files. The same data splits and preprocessing pipeline are used for all compared models and ablation variants. During training, the input image size is set to 224×224 . The data augmentation pipeline includes bicubic interpolation, color jitter, RandAugment, random erasing, Mixup, and CutMix. The color jitter factor is set to 0.4, the RandAugment policy is `rand-m9-mstd0.5-inc1`, and the random erasing probability is set to 0.25. The alpha values of Mixup and CutMix are set to 0.8 and 1.0, respectively. For validation and testing, images are resized and center-cropped to 224×224 , followed by tensor conversion and normalization using the ImageNet mean and standard deviation.

We adopt vHeat-Tiny as the base backbone. Unless otherwise specified, all models are trained from scratch with random initialization and without pretrained weights. The original vHeat, PHR-only, FDR-only, and PHR+FDR variants are constructed using the same backbone, data splits, training settings, and evaluation protocol. They

differ only in whether the corresponding components are enabled. The specific stage-wise PHR routing ratios and FDR frequency-retention settings are reported in the ablation study.

All models are trained for 300 epochs with a batch size of 128. We use the AdamW optimizer with an initial learning rate of 5×10^{-4} , a weight decay of 0.08, and a cosine learning rate schedule. The warm-up period lasts for 20 epochs, with an initial warm-up learning rate of 5×10^{-7} . The minimum learning rate is set to 5×10^{-6} . Gradient clipping is applied with a maximum norm of 5.0, label smoothing is set to 0.1, and automatic mixed precision is enabled during training.

Unless otherwise specified, the default training random seed is set to 0. To examine whether the main observations depend on a single random seed, we additionally repeat the key variants on Pollen149 using seeds 1 and 2. These variants include the original vHeat, PHR-only, FDR-only, and the final PHR+FDR configuration. We report the mean and standard deviation over the three training seeds. For the two public pollen datasets, we also repeat the original vHeat and the final PHR+FDR configuration using three training seeds. The public dataset split seed is fixed to 42 and remains unchanged across different training seeds.

During training, the best checkpoint is selected according to validation Top-1 accuracy and is subsequently evaluated on the test set. The HCO redundancy diagnostics in Section 4.2 are conducted on the held-out validation split, whereas the final recognition performance in the main comparison tables is reported on the test split. This protocol separates redundancy diagnosis and model selection from final performance evaluation. All experiments are implemented using PyTorch 2.1.0, and each training run is conducted on a single NVIDIA RTX 3090 GPU.

For model complexity and practical inference efficiency, we use a unified profiling script. The number of parameters is computed from all model parameters, and FLOPs are calculated using `thop` with a random input tensor of size $1 \times 3 \times 224 \times 224$. Inference speed is measured using random input tensors of size $B \times 3 \times 224 \times 224$, where $B = 256$. All models are evaluated using the same hardware, batch size, and input resolution. We perform 50 warm-up iterations followed by 200 timed iterations. Throughput is reported in images per second, and latency denotes the average inference time per image under batched inference.

Top-1 accuracy is used to evaluate classification performance. On Pollen149, we additionally report the number of parameters, FLOPs, throughput, and latency to evaluate model efficiency. Unless otherwise specified, the main experimental tables report results obtained with the default training seed. The additional stability analyses report the mean and standard deviation over three seeds for the key ablation variants and for the final model on the public datasets. The New Zealand Pollen Dataset and CPD-1 are used as additional validation datasets, where recognition accuracy and model complexity are mainly reported to analyze the behavior of the proposed method under different data distributions.

4.2 Redundancy Analysis

This section investigates whether HCO in vHeat contains exploitable redundancy and whether such redundancy is distributed non-uniformly across network blocks and

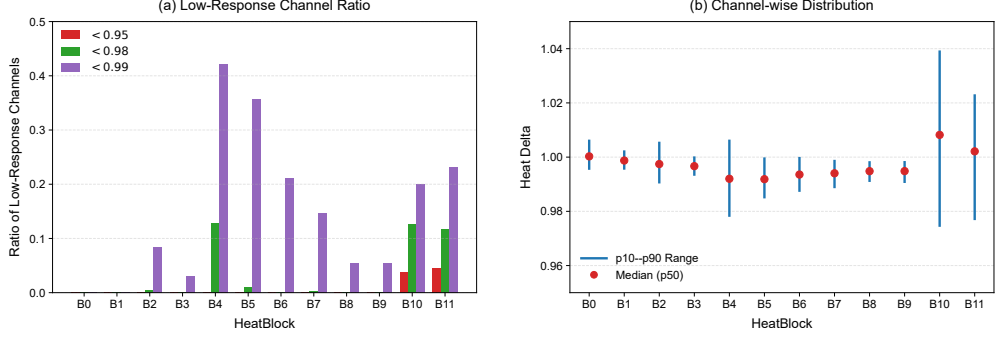


Fig. 2 Channel-path redundancy analysis of vHeat. (a) Proportions of low-response channels under different heat-delta thresholds. (b) Distribution intervals of channel-wise heat delta using p10, p50, and p90 statistics.

stages. We analyze HCO from two perspectives: the channel computation path and the DCT-domain frequency response. Response statistics are first used to identify potentially redundant regions, followed by masking experiments to examine whether these regions exhibit functional tolerance to suppression. The resulting analyses provide empirical motivation for the designs of PHR and FDR. The specific channel routing ratios and frequency-retention settings are further determined through the ablation study in Section 4.3 according to the accuracy–efficiency trade-off.

4.2.1 Channel-path Redundancy Analysis of vHeat

To examine whether HCO contains exploitable redundancy in the channel computation path, we first conduct a channel-wise response analysis on the Pollen149 validation set. A trained vHeat model is evaluated while recording the feature responses before and after the Heat2D module in each HeatBlock. Let the input and output of Heat2D in a given HeatBlock be denoted as $X \in \mathbb{R}^{B \times C \times H \times W}$ and $Y \in \mathbb{R}^{B \times C \times H \times W}$, respectively. The relative Heat2D-induced variation of the c -th channel is defined as

$$\Delta_c = \frac{\|Y_c - X_c\|_2}{\|X_c\|_2 + \varepsilon}, \quad (15)$$

where $\|\cdot\|_2$ denotes the L_2 norm over the spatial dimensions, and ε is a small constant used to avoid division by zero. This metric characterizes the relative response variation induced by Heat2D for each channel. Based on the channel-wise Δ_c , we calculate the proportions of low-response channels under different thresholds and summarize the p10, p50, and p90 response statistics within each block, as shown in Fig. 2.

Figure 2(a) shows that the proportion of low- Δ_c channels varies substantially across HeatBlocks. Several middle-to-late blocks contain a noticeable proportion of relatively low-response channels. For example, under the threshold of 0.99, the proportions reach 42.19% and 35.68% in block_4 and block_5, respectively. Low- Δ_c channels are also observed in several deeper blocks. These results indicate that Heat2D does not affect all channels uniformly across network positions.

Table 1 Performance (%) after masking low-response channels in different blocks. The baseline performance is 96.20%. Values in parentheses denote the drop compared with the baseline.

Masked Blocks	Lowest 10%	Lowest 20%	Lowest 30%
<code>block_4</code>	96.09 (-0.11)	95.66 (-0.54)	95.32 (-0.88)
<code>block_4~5</code>	95.92 (-0.28)	96.09 (-0.11)	95.81 (-0.39)
<code>block_4~7</code>	95.92 (-0.28)	95.86 (-0.34)	95.58 (-0.62)

Figure 2(b) further presents the distribution of channel responses within individual blocks. Here, p50 denotes the median response, while the p10–p90 interval characterizes the main response range across channels. Several middle-to-late blocks exhibit clear response dispersion, indicating the coexistence of relatively low- and high-response channels within the same HeatBlock. Overall, the Heat2D-induced variation is non-uniform across both blocks and channels, providing evidence of potential optimization space in the channel computation path.

4.2.2 Mask-based Verification of Channel-path Redundancy

Response statistics alone are insufficient to determine whether low-response channels correspond to functionally exploitable redundancy. We therefore perform a channel masking experiment to verify the functional contribution of the low- Δ_c channels identified above. Specifically, channels within selected HeatBlocks are ranked according to their heat-delta values, and the lowest 10%, 20%, or 30% are selected as masking targets. During inference, the outputs of these channels are set to zero, and recognition performance is re-evaluated on the validation set.

The results are reported in Table 1. Under small or moderate masking ratios, suppressing part of the low-response channels causes only limited performance degradation. For example, masking the lowest 10% of channels in `block_4` decreases Top-1 accuracy by only 0.11 percentage points. Masking the lowest 20% across `block_4~5` results in the same 0.11 percentage-point decrease. Moderate masking over a wider range of blocks also retains relatively stable recognition performance.

Performance degradation becomes more noticeable as the masking strength increases. For example, masking the lowest 30% of channels in `block_4` decreases accuracy by 0.88 percentage points. This result indicates that low-response channels are not entirely dispensable. Instead, they contain both a subset with higher tolerance to suppression and channels that still contribute useful discriminative information.

Combining the response statistics and masking results, the complete Heat2D computation exhibits exploitable optimization space along the channel dimension. The limited performance degradation under moderate suppression of selected low-response channels provides experimental support for reducing redundant full-channel Heat2D computation through PHR.

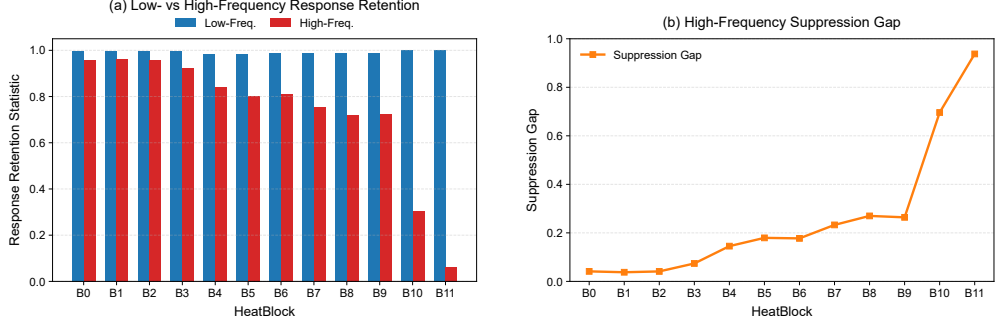


Fig. 3 Frequency-response redundancy analysis of vHeat. (a) Frequency-response retention statistics for low- and high-frequency components across different blocks. (b) Suppression gap between low- and high-frequency responses.

4.2.3 Frequency-response Redundancy Analysis of vHeat

To examine the utilization of different frequency components in HCO, we further perform frequency-response statistics on the Pollen149 validation set. A trained vHeat model is evaluated with the frequency statistics module enabled in each HeatBlock. The module records the retention behavior of low- and high-frequency responses during Heat2D processing. These statistics are used as diagnostic indicators to characterize the relative frequency-response behavior across different blocks and frequency regions. We further compute the suppression gap between low- and high-frequency responses to examine the frequency selectivity of HCO. The results are visualized in Fig. 3.

Figure 3(a) shows that HCO does not utilize different frequency responses uniformly. Low-frequency responses maintain relatively high retention statistics across most blocks, whereas high-frequency responses become increasingly suppressed from the middle stages onward. This trend is particularly evident in deeper blocks, where the high-frequency retention statistics decrease to 0.3039 and 0.0625 in block_10 and block_11, respectively.

The suppression gap shown in Fig. 3(b) further reflects this non-uniform behavior. The gap between low- and high-frequency responses is relatively small in shallow blocks but increases substantially in the middle-to-late stages. This pattern indicates that HCO exhibits stronger frequency selectivity at greater network depths.

These observations show that the frequency-domain modeling of HCO is non-uniform across frequency regions and network stages. In particular, some higher-frequency responses are already strongly suppressed in the middle-to-late stages, suggesting potential room for additional controlled frequency constraints. This observation provides frequency-domain diagnostic evidence for the design of FDR.

4.2.4 Functional Verification of Frequency-response Redundancy

To determine whether the observed frequency-response non-uniformity corresponds to functionally exploitable redundancy, we perform a frequency-response masking experiment. A low-frequency retention mask is applied to selected HeatBlocks during

Table 2 Performance (%) under frequency-response masking on different blocks. The baseline performance is 96.20%. Values in parentheses indicate the absolute performance drop relative to the baseline.

Masked Blocks	Retention Ratio 0.9	Retention Ratio 0.7	Retention Ratio 0.5
<code>block_11</code>	96.01 (-0.19)	95.90 (-0.30)	96.01 (-0.19)
<code>block_10~11</code>	95.96 (-0.24)	95.86 (-0.34)	95.58 (-0.62)
<code>block_8~11</code>	96.09 (-0.11)	95.79 (-0.41)	95.12 (-1.08)

inference, and model performance is re-evaluated on the validation set. Because this experiment directly modifies the frequency responses used for subsequent IDCT reconstruction, it provides a functional test of whether the identified frequency regions can be constrained with limited performance loss.

The results are shown in Table 2. Moderate frequency constraints applied to deeper HeatBlocks cause only limited degradation in recognition performance. For example, when masking is applied only to `block_11`, the accuracy decreases by no more than 0.30 percentage points across the evaluated retention settings. When the masking range is extended to `block_10~11`, moderate frequency constraints still result in relatively small performance losses.

Stronger or more widely applied frequency suppression leads to more noticeable degradation. For example, when a retention ratio of 0.5 is applied to `block_8~11`, the accuracy decreases by 1.08 percentage points. This indicates that the suppressed frequency responses are not entirely redundant and that some of them still contain useful discriminative information.

The frequency-response statistics and masking experiments together show that part of the middle-to-late-stage frequency responses can tolerate controlled constraints, whereas overly strong or extensive suppression may remove useful information. These results support the use of FDR as a controlled frequency-domain regularization mechanism rather than aggressive frequency removal.

Combining the channel-path and frequency-response analyses, HCO exhibits non-uniform and exploitable optimization space in both dimensions. Some relatively low-response channels tolerate moderate suppression, while part of the middle-to-late-stage frequency responses can also be constrained with limited performance degradation. These diagnostic and functional verification results provide empirical support for the designs of PHR and FDR, respectively.

4.3 Ablation Study on PHR and FDR

To verify the effectiveness of Partial Heat Routing (PHR) and Frequency-domain Regularization (FDR), and to determine the final stage-wise deployment configuration, we conduct systematic ablation experiments, as shown in Table 3.

It should be noted that the redundancy analysis in Section 4.2 is mainly used to identify potential optimization space in the channel computation path and frequency responses of HCO, and to guide the design of candidate configurations. The final channel routing ratios and frequency-retention ratios are not directly calculated

from the redundancy statistics. Specifically, the channel-path analysis shows that several HeatBlocks tolerate moderate channel suppression, while the frequency-response analysis indicates that part of the middle-to-late-stage frequency responses can tolerate controlled constraints. Based on these observations, we further compare uniform, stage-wise, aggressive, and block-wise configurations. The specific combinations of r_s and ρ_s are finally selected according to the accuracy–efficiency trade-off reflected by Top-1 accuracy, parameter count, FLOPs, measured throughput, and latency.

The original vHeat achieves a Top-1 accuracy of 96.20% for a 224×224 input, with 30.20M parameters and 4.94G FLOPs. It is used as the baseline model in this set of experiments. The comparison among different configurations shows that PHR and FDR play different roles in HCO optimization. PHR mainly reduces structural computational cost by adjusting the channel computation path, whereas FDR mainly regulates the frequency-domain representation through controlled frequency-response constraints.

Using PHR alone clearly reduces model complexity, but also introduces a certain accuracy drop. Uniform PHR reduces the number of parameters and FLOPs to 24.93M and 4.01G, respectively, while achieving a Top-1 accuracy of 95.66%. In comparison, stage-wise PHR achieves 95.79% Top-1 accuracy with a throughput of 1020.78 images/s under a similar computational budget. These results indicate that applying the same routing ratio to all stages is not necessarily optimal. Different network stages exhibit different tolerance to partial channel routing, and stage-wise deployment provides a better balance between structural computational cost and recognition performance.

Unlike PHR, FDR is not primarily designed for structural acceleration. With unchanged parameters and FLOPs, uniform FDR achieves a Top-1 accuracy of 96.35%, while stage-wise FDR further improves it to 96.48%. These results indicate that moderate frequency-response constraints do not necessarily damage useful information and can act as an effective frequency-domain regularization mechanism. Meanwhile, the FDR-only configurations do not improve throughput. This further suggests that the main role of FDR is to regulate frequency-domain representation rather than directly reduce structural computation. The observation is also consistent with the frequency-response redundancy analysis in Section 4.2.

When PHR and FDR are jointly applied, the configuration strategy has a clear influence on the final performance. Uniform PHR+FDR improves inference efficiency but reduces the Top-1 accuracy to 95.45%, indicating that uniformly applying channel routing and frequency constraints across all stages may cause excessive information loss. The aggressive configuration achieves the highest throughput of 1039.14 images/s, but also introduces a larger accuracy drop.

In comparison, the stage-wise PHR+FDR configuration reduces the number of parameters from 30.20M to 25.69M and FLOPs from 4.94G to 3.97G. At the same time, throughput increases from 924.15 images/s to 1015.27 images/s, while latency decreases from 1.0821 ms to 0.9850 ms. The corresponding Top-1 accuracy is 96.01%, which is only 0.19 percentage points lower than that of the original vHeat. We therefore adopt this configuration as the final model, denoted as Ours in Table 3. This choice does not target the highest standalone accuracy or the maximum inference speed.

Table 3 Ablation study of Partial Heat Routing (PHR) and Frequency-domain Regularization (FDR). FLOPs are reported for a single 224×224 input image. Throughput and latency are measured under the unified profiling protocol, and latency denotes the average inference time per image under batched inference.

ID	Method	Config.	Top-1 (%)	Δ Acc.	Params	FLOPs	Thr.	Lat.
1	Original vHeat	–	96.20	0.00	30.20M	4.94G	924.15	1.0821
2	PHR-only	Uniform PHR	95.66	-0.54	24.93M	4.01G	1000.12	0.9999
3	PHR-only	Stage-wise PHR	95.79	-0.41	25.69M	3.97G	1020.78	0.9796
4	FDR-only	Uniform FDR	96.35	+0.15	30.20M	4.94G	897.90	1.1137
5	FDR-only	Stage-wise FDR	96.48	+0.28	30.20M	4.94G	910.66	1.0981
6	PHR+FDR	Uniform	95.45	-0.75	24.93M	4.01G	985.30	1.0149
7	PHR+FDR, Ours	Stage-wise	96.01	-0.19	25.69M	3.97G	1015.27	0.9850
8	PHR+FDR	Aggressive	95.75	-0.45	23.72M	3.79G	1039.14	0.9623
9	PHR+FDR	Stage-2 Block-wise	96.33	+0.13	28.92M	4.68G	903.25	1.1071

Note: Thr. and Lat. denote throughput (img/s) and latency (ms/img), respectively. Uniform PHR uses $r = 0.625$, while stage-wise PHR uses $r = [0.5/0.5/0.625/0.75]$. Uniform FDR uses $\rho = 0.8$, while stage-wise FDR uses $\rho = [1.0/1.0/0.8/0.6]$. The aggressive setting uses $r = [0.5/0.5/0.5/0.5]$ and $\rho = [1.0/1.0/0.7/0.5]$. Here, ρ denotes the frequency-retention ratio.

Instead, it jointly considers recognition performance, model complexity, and practical inference efficiency.

We further include a Stage-2 Block-wise configuration to examine whether channel routing should be concentrated on the middle network positions indicated by the redundancy analysis. This configuration achieves a Top-1 accuracy of 96.33%, exceeding that of the original vHeat. The result indicates that the middle HeatBlocks exhibit relatively high tolerance to partial channel routing, which is consistent with the channel-path redundancy analysis in Section 4.2.

However, the Stage-2 Block-wise configuration still requires 28.92M parameters and 4.68G FLOPs, and its measured throughput is only 903.25 images/s. This is lower than that of the final stage-wise PHR+FDR configuration. These results show that redundancy tolerance alone is insufficient to determine the final deployment positions. Practical acceleration also depends on the computational contribution of different stages. Earlier stages operate at higher spatial resolutions and therefore have a larger influence on Heat2D computational cost and measured inference speed. The final configuration is therefore not determined solely by the locations with stronger redundancy. It also considers stage-wise computational cost, recognition performance, and practical inference efficiency.

Overall, PHR and FDR optimize HCO from the channel-path and frequency-response dimensions, respectively. The final stage-wise PHR+FDR configuration reduces the number of parameters and FLOPs by 14.9% and 19.6%, respectively. In practical inference, throughput improves by approximately 9.9%, while latency decreases by approximately 9.0%. Under the default training seed, the Top-1 accuracy decreases by only 0.19 percentage points. These results indicate a favorable accuracy–efficiency trade-off for vHeat-based pollen recognition.

To further examine whether the main observations depend on a single training seed, we repeat the key configurations using three random seeds, as shown in Table 4. Table 3 provides the complete comparison of accuracy, model complexity, and practical inference efficiency under the default training seed. The three-seed experiments serve

Table 4 Three-seed stability analysis of key ablation variants on Pollen149. Results are reported as Top-1 accuracy (%).

Method	Top-1 Acc. (%)
Original vHeat	95.93 ± 0.31
PHR-only	95.70 ± 0.10
FDR-only	96.16 ± 0.28
PHR+FDR (Ours)	95.90 ± 0.11

Note: The mean and standard deviation are computed over three training seeds using the sample standard deviation.

as a supplementary stability analysis to examine whether the recognition-performance trends of the key configurations remain consistent across different training seeds.

The original vHeat achieves a Top-1 accuracy of $95.93 \pm 0.31\%$, while the final PHR+FDR configuration achieves $95.90 \pm 0.11\%$. Their mean recognition performance is therefore comparable, indicating that the final model maintains vHeat-level recognition accuracy across different training seeds. PHR-only achieves $95.70 \pm 0.10\%$, suggesting that channel routing alone introduces a certain accuracy cost. FDR-only achieves $96.16 \pm 0.28\%$. Its mean accuracy is higher than that of the original vHeat, which is consistent with the positive accuracy trend observed for FDR under the default seed. Combined with the complexity and inference-efficiency results in Table 3, the final PHR+FDR configuration reduces parameters, FLOPs, and inference cost while maintaining recognition performance comparable to the original vHeat.

4.4 Further Analysis

4.4.1 Effect of FDR Frequency-retention Ratio

To further analyze the influence of the frequency-retention strength in Frequency-domain Regularization, we evaluate different frequency-retention ratios, as shown in Table 5. The setting with $\rho = 1.0$ applies no frequency-response constraint and corresponds to the original vHeat baseline, which achieves a Top-1 accuracy of 96.20%.

The results show that reducing the frequency-retention ratio does not lead to a monotonic improvement in recognition performance. With a uniform setting of $\rho = 0.8$, the Top-1 accuracy increases to 96.35%. The stage-wise configuration [1.0/1.0/0.8/0.6] further improves the accuracy to 96.48%, which is the best result in this group of experiments. This suggests that different network stages have different sensitivities to frequency-response constraints, and that stage-wise deployment is more suitable than applying the same constraint to all stages.

As the frequency constraint becomes stronger, recognition performance starts to decrease. For example, the model still achieves 96.31% Top-1 accuracy with $\rho = 0.6$, whereas reducing ρ to 0.5 decreases the accuracy to 95.94%. Therefore, the objective of FDR is not to suppress as many frequency responses as possible, but to apply a

Table 5 Effect of different FDR frequency-retention ratios.

Retention Ratio	Top-1 Acc. (%)	Change
1.0 (Baseline)	96.20	0.00
0.8 (Uniform)	96.35	+0.15
0.7 (Uniform)	96.20	0.00
0.6 (Uniform)	96.31	+0.11
0.5 (Uniform)	95.94	-0.26
Stage-wise [1.0/1.0/0.8/0.6]	96.48	+0.28

moderate frequency-domain constraint while preserving useful discriminative information. Based on this analysis, we adopt the stage-wise configuration [1.0/1.0/0.8/0.6] as the default FDR setting.

4.4.2 Comparison with General Pruning Criteria

To further examine the benefit of redundancy-aware optimization tailored to the structural characteristics of HCO, we conduct a supplementary comparison with two general pruning criteria: Magnitude Pruning (MP) [48] and Taylor Pruning [49]. The results are shown in Table 6. Here, “Pruned Acc.” denotes the accuracy immediately after pruning without fine-tuning, while “Fine-tuned Acc.” denotes the best accuracy obtained after 20 epochs of post-pruning fine-tuning.

Directly applying the two pruning methods to vHeat causes substantial performance degradation. Magnitude Pruning achieves a Top-1 accuracy of 66.20% immediately after pruning and recovers to 94.13% after 20 epochs of fine-tuning. Taylor Pruning achieves 80.01% after pruning and further recovers to 94.67% after fine-tuning. Under the considered experimental setting, Taylor Pruning therefore preserves recognition performance better than Magnitude Pruning. However, the fine-tuned accuracies of both pruning baselines remain lower than those of the original vHeat and the final PHR+FDR configuration.

Unlike these pruning criteria, which determine pruning decisions based on weight magnitude or gradient sensitivity, the proposed method performs targeted optimization according to the channel computation path and frequency-response characteristics of HCO. The final PHR+FDR configuration maintains a Top-1 accuracy of 96.01% while reducing model complexity and practical inference cost. Under the two general pruning baselines considered here, the proposed HCO-oriented optimization therefore provides better preservation of recognition performance.

4.4.3 Additional Validation on Public Pollen Datasets

To further evaluate the applicability of the proposed method under different data sources, we conduct additional experiments on the New Zealand Pollen Dataset and CPD-1. Compared with the primary Pollen149 dataset, these two public datasets differ in sample scale, category composition, imaging modality, and data distribution. This section therefore examines whether the proposed HCO-oriented optimization

Table 6 Comparison with general pruning criteria. MP and Taylor models are fine-tuned for 20 epochs after pruning.

Method	Pruned Acc. (%)	Fine-tuned Acc. (%)
Original vHeat	–	96.20
Magnitude Pruning	66.20	94.13
Taylor Pruning	80.01	94.67
Ours	–	96.01

Table 7 Additional validation results on the New Zealand Pollen Dataset and CPD-1. FLOPs are reported for a single 224×224 input image.

Dataset	Method	Top-1 (%)	Params (M)	FLOPs (G)
New Zealand	vHeat	98.60	30.10	4.94
	PHR-only	98.88	25.61	3.97
	FDR-only	99.01	30.10	4.94
	Ours	98.73	25.61	3.97
CPD-1	vHeat	97.64	30.10	4.94
	PHR-only	97.39	25.59	3.97
	FDR-only	98.26	30.10	4.94
	Ours	98.01	25.59	3.97

can maintain a favorable accuracy–complexity trade-off under different pollen data conditions. The default-seed results are reported in Table 7.

On the New Zealand Pollen Dataset, the original vHeat achieves a Top-1 accuracy of 98.60%. PHR-only reaches 98.88% while reducing the number of parameters and FLOPs. FDR-only achieves the highest accuracy of 99.01%. The final PHR+FDR configuration obtains 98.73% Top-1 accuracy. Although its accuracy is lower than that of FDR-only, it remains higher than the original vHeat while preserving the complexity reduction introduced by PHR.

On CPD-1, PHR-only slightly decreases the Top-1 accuracy from 97.64% to 97.39%, whereas FDR-only improves it to 98.26%. The final PHR+FDR configuration achieves 98.01% while maintaining the reduced parameter count and FLOPs of the PHR-based structure. The results on both public datasets show that the final configuration is not designed to maximize standalone accuracy. Instead, it aims to maintain or improve the recognition performance of the original vHeat under lower model complexity.

To further examine the stability of the final configuration on public datasets, we repeat the original vHeat and Ours using three training seeds, as shown in Table 8. The default-seed experiments provide component-wise comparisons, while the three-seed experiments serve as a supplementary stability analysis of the final configuration under different pollen datasets.

On the New Zealand Pollen Dataset, vHeat and Ours achieve Top-1 accuracies of $99.00 \pm 0.35\%$ and $99.04 \pm 0.27\%$, respectively. On CPD-1, the corresponding results are

Table 8 Three-seed external stability analysis on public pollen datasets. Results are reported as Top-1 accuracy (%).

Dataset	Method	Top-1 Acc. (%)
New Zealand	vHeat	99.00 ± 0.35
	Ours	99.04 ± 0.27
CPD-1	vHeat	98.31 ± 0.60
	Ours	98.44 ± 0.51

Note: The mean and standard deviation are computed over three training seeds using the sample standard deviation.

$98.31 \pm 0.60\%$ and $98.44 \pm 0.51\%$. Overall, the final model maintains average recognition performance comparable to the original vHeat on both public datasets while retaining its advantage in model complexity.

4.4.4 Visualization Analysis

To further analyze how PHR and FDR influence model representation behavior, we provide two types of qualitative visualization: frequency-response visualization and Eigen-CAM activation maps. The former illustrates the DCT-domain frequency-response patterns of different model variants across network stages, while the latter shows the spatial regions attended to during classification. These visualizations provide qualitative evidence complementary to the quantitative results.

Figure 4 presents the mean frequency-response retention statistics of different model variants across stages. In shallow stages, all variants generally exhibit high frequency-response retention. This behavior is consistent with the conservative frequency constraints used by FDR in early stages, which avoid imposing excessive restrictions on the formation of early visual features.

In the middle and deep stages, the effect of FDR becomes more evident. Both FDR-only and PHR+FDR exhibit clearer frequency selectivity, with low- and mid-frequency regions being relatively preserved while part of the high-frequency responses are suppressed. In contrast, PHR-only does not exhibit a similar frequency-selective pattern. This is consistent with the design of PHR, which mainly modifies the channel computation path rather than directly imposing frequency-domain constraints. The PHR+FDR variant maintains a frequency-response pattern similar to FDR-only in deeper stages, indicating that the frequency-domain constraint remains effective after partial channel routing.

Figure 5 further presents the Eigen-CAM results of different models. In the representative samples, the original vHeat generally attends to the pollen body, although some activations also appear in background regions, boundary areas, or other structures. After introducing FDR, the responses in some samples are more concentrated on morphology-related regions, including contours, grooves, pores, surface textures, and local protrusions. These observations suggest that FDR may encourage the model

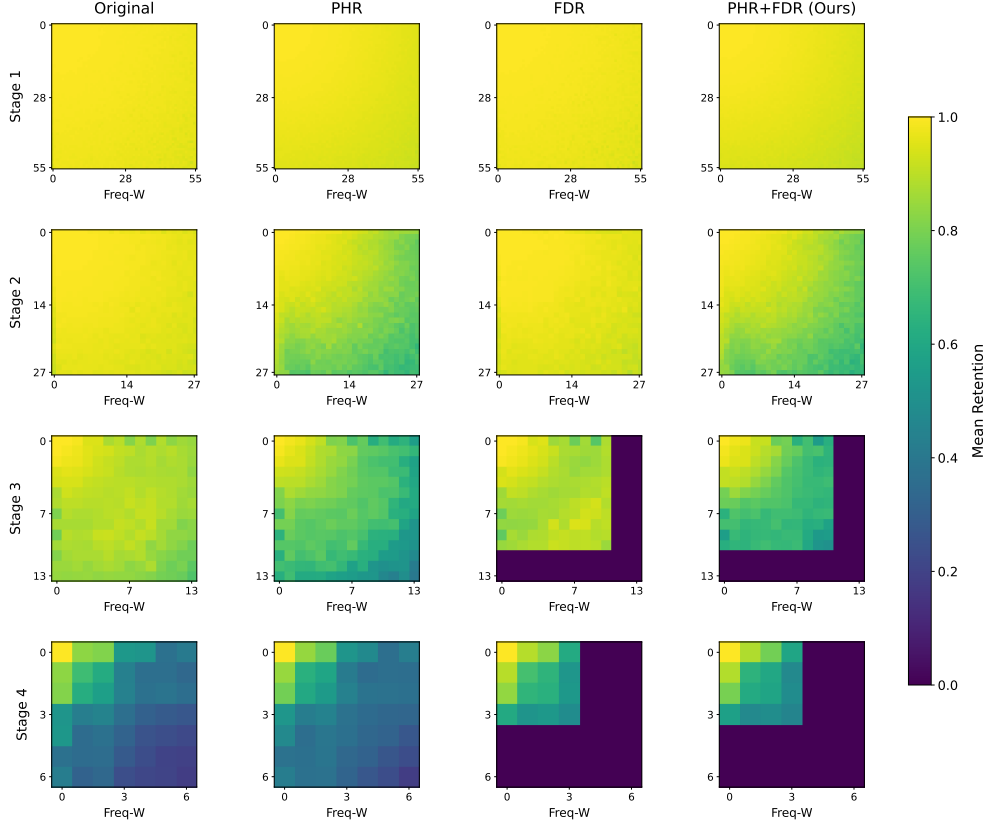


Fig. 4 Visualization of mean frequency-response retention statistics across different stages and model variants. Brighter regions indicate higher frequency-response retention statistics, while darker regions indicate stronger response suppression.

to attend more strongly to morphology-related discriminative regions in the visualized examples.

For PHR+FDR, the Eigen-CAM results show that meaningful spatial responses are still preserved after partial channel routing. The activation regions generally remain around the pollen body and its local morphological structures. Together with the frequency-response visualization, these results show that FDR produces clearer frequency selectivity in the middle and deep stages, while PHR+FDR maintains reasonable spatial discriminative responses under a more efficient HCO configuration.

4.5 Comparison with Representative Backbones

To provide a broader reference for recognition performance and efficiency, we compare the proposed method with representative visual backbones on Pollen149, including CNN-based, Transformer-based, lightweight, and heat-based architectures. The results are reported in Table 9. This comparison illustrates the overall accuracy-efficiency

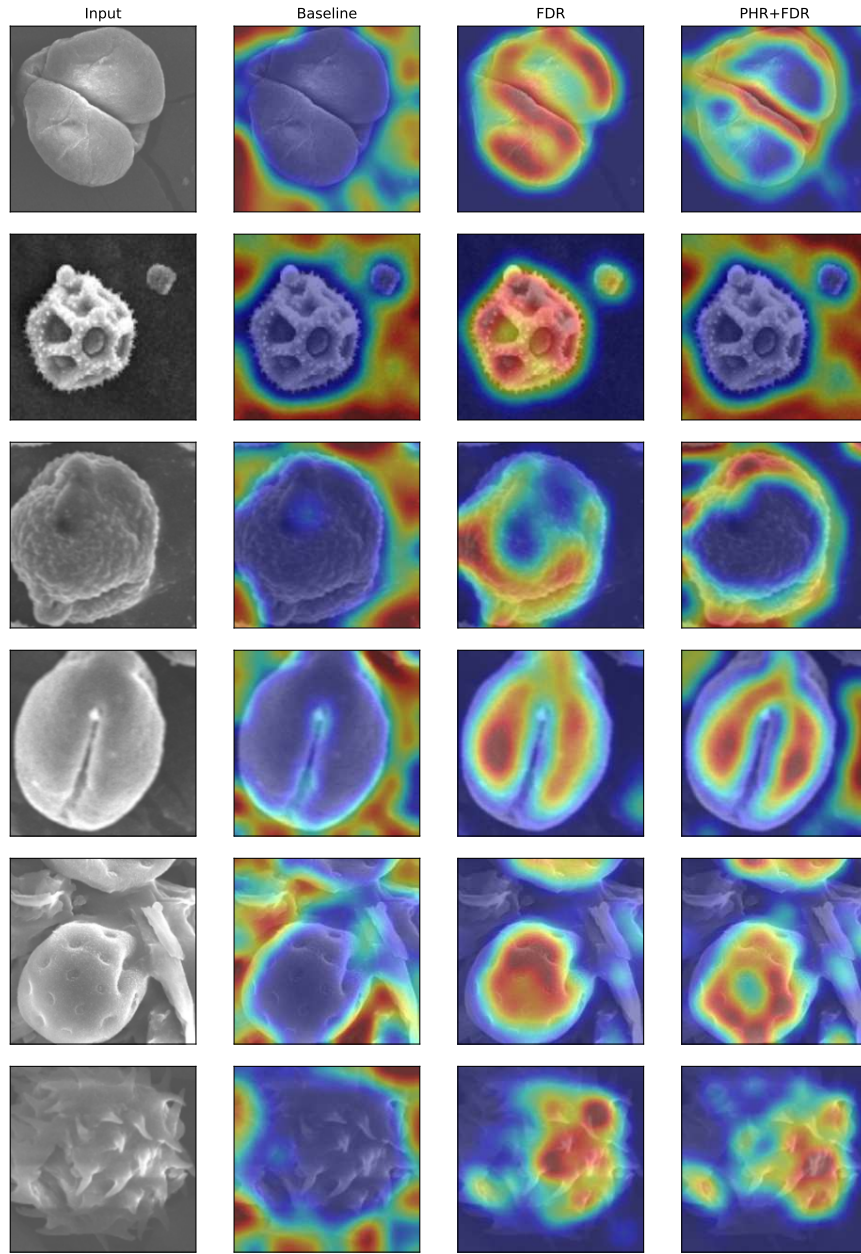


Fig. 5 Eigen-CAM visualization of the original vHeat, FDR-only, and PHR+FDR models. The visualizations provide qualitative evidence of the spatial regions used for classification.

Table 9 Comparison with representative backbones on Pollen149.

Method	Top-1 (%)	Params (M)	FLOPs (G)	Throughput (img/s)	Latency (ms)
ResNet [50]	87.30	23.81	4.11	1563.09	0.6398
Swin Transformer [51]	95.60	27.63	4.51	832.89	1.2006
ConvNeXt [52]	86.20	27.93	4.47	944.36	1.0589
RepViT [30]	93.87	6.47	1.12	2585.70	0.3867
MogaNet [53]	85.10	4.98	1.09	1224.81	0.8165
vHeat [10]	96.20	30.20	4.94	924.15	1.0821
Ours	96.01	25.69	3.97	1015.27	0.9850

characteristics of different backbone families. The main focus of this work, however, remains on whether redundant HCO computation can be reduced while largely maintaining the recognition performance of the original vHeat backbone.

The compared backbone families exhibit different accuracy–efficiency characteristics. RepViT achieves the highest throughput and the lowest latency, while MogaNet has the smallest parameter count and FLOPs. ResNet and ConvNeXt also provide relatively high inference throughput, but their recognition accuracies are lower than those of vHeat and the proposed method. Swin Transformer achieves the highest Top-1 accuracy among the non-vHeat backbones, although it remains below the two vHeat-based models. Overall, lightweight and convolution-based models provide clear advantages in raw computational efficiency, whereas vHeat achieves higher recognition accuracy on Pollen149.

The comparison between the original vHeat and Ours more directly reflects the effect of the proposed optimization. Ours reduces the number of parameters from 30.20M to 25.69M and FLOPs from 4.94G to 3.97G. At the same time, throughput increases from 924.15 images/s to 1015.27 images/s. The Top-1 accuracy remains close to that of the original vHeat, decreasing only from 96.20% to 96.01%. These results show that the proposed redundancy-aware optimization improves the accuracy–efficiency trade-off of vHeat. The proposed method should therefore be interpreted as an HCO-oriented optimization strategy for vHeat rather than a general-purpose lightweight backbone.

4.6 Limitations and Future Work

Although the proposed method improves the accuracy–efficiency trade-off of vHeat for fine-grained pollen recognition, several limitations should be acknowledged. First, the current experiments are conducted on pollen image datasets, including the primary Pollen149 dataset and two public pollen datasets. The conclusions are therefore most directly supported for pollen recognition scenarios. When applying PHR and FDR to other microscopic image recognition tasks or more general visual recognition tasks, the stage-wise routing ratios and frequency-retention ratios may need to be re-tuned according to the data distribution, imaging modality, and backbone configuration.

Second, the practical inference acceleration is smaller than the theoretical FLOPs reduction. In the main experiment, the final configuration reduces FLOPs by 19.6%, whereas measured throughput improves by approximately 9.9% and latency decreases by approximately 9.0%. Practical inference speed depends not only on theoretical

computation but also on factors such as memory access, operator implementation, GPU utilization, and branch execution. The proposed method should therefore be interpreted as an HCO-oriented accuracy–efficiency optimization rather than a purely hardware-optimized acceleration scheme. Future work may further improve the execution efficiency of partial Heat routing and frequency constraints on different hardware platforms.

Third, although FDR is designed as a controlled frequency-domain regularization mechanism rather than aggressive frequency pruning, frequency constraints may still suppress useful high-frequency microscopic details. For pollen recognition, fine-grained morphological structures such as surface ornamentation, pores, grooves, and local protrusions may provide important discriminative cues. Overly strong or improperly configured frequency constraints may therefore weaken useful morphological information. In this work, this risk is mitigated by adopting conservative frequency-retention settings in early stages and selecting the final configuration through ablation experiments. Future work may explore more flexible frequency constraints to better balance efficiency optimization and the preservation of fine-grained high-frequency information.

5 Conclusion

This paper investigates the accuracy–efficiency trade-off of the heat-conduction-based vision backbone vHeat for fine-grained pollen recognition from the perspective of redundancy within the Heat Conduction Operator (HCO). Through channel-response statistics, channel masking, frequency-response statistics, and frequency masking experiments, we identify non-uniform and potentially exploitable optimization space in both the channel computation path and the frequency-response dimensions of HCO. Based on these diagnostic observations, we propose an HCO-oriented redundancy-aware optimization framework consisting of Partial Heat Routing (PHR), Frequency-domain Regularization (FDR), and a stage-wise deployment strategy.

PHR reduces full-channel Heat2D computation by reallocating part of the channels from the complete Heat2D branch to a lightweight local branch while preserving the original output channel dimensionality. FDR introduces a controlled frequency-retention constraint in the DCT domain to regulate the use of different frequency responses within HCO. Since different network stages differ in spatial resolution, channel dimension, computational contribution, and representation level, a stage-wise deployment strategy is further used to coordinate the channel routing ratios of PHR and the frequency-retention ratios of FDR. The redundancy analysis guides the design of candidate configurations, while the final configuration is selected through ablation experiments by jointly considering recognition performance, model complexity, and practical inference efficiency.

On Pollen149, the final PHR+FDR configuration reduces the number of parameters from 30.20M to 25.69M and FLOPs from 4.94G to 3.97G. Under the default training seed, its Top-1 accuracy is only 0.19 percentage points lower than that of the original vHeat. At the same time, throughput increases from 924.15 images/s to 1015.27 images/s, while latency decreases from 1.0821 ms to 0.9850 ms under the

unified profiling protocol. The three-seed analysis further shows that the final model maintains average recognition performance comparable to the original vHeat. Additional validation on two public pollen datasets also indicates that the proposed method can maintain a favorable accuracy–complexity trade-off under different pollen data conditions.

Overall, the results show that targeted optimization of the non-uniform channel-path and frequency-response behavior within HCO can reduce model complexity and practical inference cost while largely preserving the recognition capability of vHeat. The proposed method should therefore be interpreted as an HCO-oriented accuracy–efficiency optimization framework for vHeat rather than as a conventional general-purpose pruning or lightweight backbone design. The current experiments mainly focus on pollen image recognition, and the applicability of the proposed strategy to other microscopic imaging tasks and broader visual recognition scenarios remains to be further investigated. Future work may explore more flexible channel routing and frequency constraints, as well as more hardware-aware implementations for efficient microscopic image recognition.

Declarations

Funding. This work was supported by Science and Technology Plan Project of Inner Mongolia Autonomous Region (2025YFHH0083, 2025YFHH0215), Inner Mongolia Natural Science Foundation Project (2025MS06039, 2023LHMS07016, 2025QN06005), Aerospace Medical Health Technology Group Co., Ltd. Medical and Industrial Integration Project (2025YK13).

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. Not applicable.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Data availability. The public datasets used in this study, including the New Zealand Pollen Dataset and CPD-1, are available from their original sources cited in the manuscript. The pollen dataset used and analyzed during the current study is not publicly available because the study is ongoing, but it can be obtained from the corresponding author upon reasonable request.

Code availability. The source code and reproducibility materials supporting this study are publicly available in the [GitHub repository](#). The repository contains the core implementation, main configuration files, data organization instructions, representative training and evaluation commands, and profiling scripts for model complexity and inference efficiency. For datasets that cannot be redistributed, the repository provides the expected data organization format and reproduction instructions rather than the restricted image data.

Author contributions. Hongyu Song contributed to conceptualization, methodology, software, validation, formal analysis, visualization, and writing—original draft. Xinru Zhang contributed to data curation, validation, and writing—review and editing. Yuan Xue contributed to experiments, visualization, and writing—review and editing. Mengfei Wang contributed to data preparation and analysis. Bao Shi supervised the study and contributed to methodology refinement, project administration, and writing—review and editing. Jingping Yang provided resources, domain guidance, and writing—review and editing. All authors read and approved the final manuscript.

References

- [1] Sofiev, M., Bergmann, K.-C. (eds.): Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts. Springer, Dordrecht (2013). <https://doi.org/10.1007/978-94-007-4881-1>
- [2] Anderegg, W.R., Abatzoglou, J.T., Anderegg, L.D., Bielory, L., Kinney, P.L., Ziska, L.: Anthropogenic climate change is worsening north american pollen seasons. *Proceedings of the National Academy of Sciences* **118**(7), 2013284118 (2021)
- [3] Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., Traidl-Hoffmann, C., Schmidt-Weber, C., Buters, J.: Automatic and online pollen monitoring. *International archives of allergy and immunology* **167**(3), 158–166 (2015)
- [4] Viertel, P., König, M.: Pattern recognition methodologies for pollen grain image classification: a survey: P. viertel, m. könig. *Machine Vision and Applications* **33**(1), 18 (2022)
- [5] Sevillano, V., Aznarte, J.L.: Improving classification of pollen grain images of the polen23e dataset through three different applications of deep learning convolutional neural networks. *PloS one* **13**(9), 0201807 (2018)
- [6] Mahbod, A., Schaefer, G., Ecker, R., Ellinger, I.: Pollen grain microscopic image classification using an ensemble of fine-tuned deep convolutional neural networks. In: *International Conference on Pattern Recognition*, pp. 344–356 (2021). Springer
- [7] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: *Proc. Int. Conf. Learn. Representations (ICLR)* (2021)
- [8] Zolfaghari, M., Sajedi, H.: Automated classification of pollen grains microscopic images using cognitive attention based on human two visual streams hypothesis. *PloS one* **19**(11), 0309674 (2024)

- [9] Zu, B., Cao, T., Li, Y., Li, J., Wang, H., Wang, Q.: Reswint: enhanced pollen image classification with parallel window transformer and coordinate attention. *The Visual Computer* **41**(7), 4975–4990 (2025)
- [10] Wang, Z., Liu, Y., Tian, Y., Liu, Y., Wang, Y., Ye, Q.: Building vision models upon heat conduction. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 9707–9717 (2025)
- [11] Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., Siniscalco, C.: Pollen13k: A large scale microscope pollen grain image dataset. In: *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2456–2460 (2020). IEEE
- [12] Rostami, M.A., Balmaki, B., Dyer, L.A., Allen, J.M., Sallam, M.F., Frontalini, F.: Efficient pollen grain classification using pre-trained convolutional neural networks: a comprehensive study. *Journal of big data* **10**(1), 151 (2023)
- [13] Yu, X., Zhao, J., Xu, Z., Wei, J., Wang, Q., Shen, F., Yang, X., Guo, Z.: Aipollen: an analytic website for pollen identification through convolutional neural networks. *Plants* **13**(22), 3118 (2024)
- [14] Konijn, T., Bijl, I., Cao, L., Verbeek, F.: Analysis of 3d urticaceae pollen classification using deep learning models. *arXiv preprint arXiv:2503.07419* (2025)
- [15] Le, G., Bao, S., Yang, D., Duan, K.: Ida: an improved dual attention module for pollen classification. *The visual computer* **40**(6), 4167–4181 (2024)
- [16] Shamrat, F.J.M., Idris, M.Y.I., Zhou, X., Khalid, M., Sharmin, S., Sharmin, Z., Ahmed, K., Moni, M.A.: Pollennet: A novel architecture for high precision pollen grain classification through deep learning and explainable ai. *Heliyon* **10**(19) (2024)
- [17] Zhang, C.-J., Liu, T., Wang, J., Zhai, D., Chen, M., Gao, Y., Yu, J., Wu, H.-Z.: Deeppollencount: a swin-transformer-yolov5-based deep learning method for pollen counting in various plant species. *Aerobiologia* **40**(3), 425–436 (2024)
- [18] Struniawski, K., Machlanska, A., Marasek-Ciolakowska, A., Konopka, A.: Automated pollen classification via subinstance recognition: A comprehensive comparison of classical and deep learning architectures. *Applied Sciences* **16**(2), 720 (2026)
- [19] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017)
- [20] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520 (2018)

- [21] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., *et al.*: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314–1324 (2019)
- [22] Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., *et al.*: Mobilenetv4: Universal models for the mobile ecosystem. In: European Conference on Computer Vision, pp. 78–96 (2024). Springer
- [23] Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856 (2018)
- [24] Ma, N., Zhang, X., Zheng, H.-T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 116–131 (2018)
- [25] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR
- [26] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1580–1589 (2020)
- [27] Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., Wang, Y.: Ghostnetv2: Enhance cheap operation with long-range attention. *Advances in Neural Information Processing Systems* **35**, 9969–9982 (2022)
- [28] Mehta, S., Rastegari, M.: Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178* (2021)
- [29] Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H., Yuan, Y.: Efficientvit: Memory efficient vision transformer with cascaded group attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14420–14430 (2023)
- [30] Wang, A., Chen, H., Lin, Z., Han, J., Ding, G.: Repvit: Revisiting mobile cnn from vit perspective. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15909–15920 (2024)
- [31] Wang, C.-Y., Liao, H.-Y.M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., Yeh, I.-H.: Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 390–391 (2020)

- [32] Huang, G., Chen, D., Li, T., Wu, F., Van Der Maaten, L., Weinberger, K.Q.: Multi-scale dense networks for resource efficient image classification. arXiv preprint arXiv:1703.09844 (2017)
- [33] Wang, Y., Huang, R., Song, S., Huang, Z., Huang, G.: Not all images are worth 16x16 words: Dynamic transformers for efficient image recognition. *Advances in neural information processing systems* **34**, 11960–11973 (2021)
- [34] Han, Y., Han, D., Liu, Z., Wang, Y., Pan, X., Pu, Y., Deng, C., Feng, J., Song, S., Huang, G.: Dynamic perceiver for efficient visual recognition. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5992–6002 (2023)
- [35] Lin, J., Rao, Y., Lu, J., Zhou, J.: Runtime neural pruning. *Advances in neural information processing systems* **30** (2017)
- [36] Meng, L., Li, H., Chen, B.-C., Lan, S., Wu, Z., Jiang, Y.-G., Lim, S.-N.: Adavit: Adaptive vision transformers for efficient image recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12309–12318 (2022)
- [37] Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S., Zhang, C.: Learning efficient convolutional networks through network slimming. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2736–2744 (2017)
- [38] Courbariaux, M., Bengio, Y., David, J.-P.: Binaryconnect: Training deep neural networks with binary weights during propagations. *Advances in neural information processing systems* **28** (2015)
- [39] Rastegari, M., Ordonez, V., Redmon, J., Farhadi, A.: Xnor-net: Imagenet classification using binary convolutional neural networks. In: *European Conference on Computer Vision*, pp. 525–542 (2016). Springer
- [40] Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
- [41] Tai, C., Xiao, T., Zhang, Y., Wang, X., et al.: Convolutional neural networks with low-rank regularization. arXiv preprint arXiv:1511.06067 (2015)
- [42] Rao, Y., Zhao, W., Zhu, Z., Lu, J., Zhou, J.: Global filter networks for image classification. *Advances in neural information processing systems* **34**, 980–993 (2021)
- [43] Li, X., Zhang, Y., Yuan, J., Lu, H., Zhu, Y.: Discrete cosin transformer: Image modeling from frequency domain. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5468–5478 (2023)

- [44] Xie, W., Song, D., Xu, C., Xu, C., Zhang, H., Wang, Y.: Learning frequency-aware dynamic network for efficient super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4308–4317 (2021)
- [45] Hu, H., Wang, P., Bi, H., Tong, B., Wang, Z., Diao, W., Chang, H., Feng, Y., Zhang, Z., Wang, Y., *et al.*: Rs-vheat: Heat conduction guided efficient remote sensing foundation model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9876–9887 (2025)
- [46] Sevillano, V., Holt, K., Aznarte, J.L.: Precise automatic classification of 46 different pollen types with convolutional neural networks. Plos one **15**(6), 0229751 (2020)
- [47] Tsiknakis, N., Savvidaki, E., Kafetzopoulos, S., Manikis, G., Vidakis, N., Marias, K., Alissandrakis, E.: Cretan Pollen Dataset v1 (CPD-1). Zenodo (2021). <https://doi.org/10.5281/zenodo.4756361> . <https://doi.org/10.5281/zenodo.4756361>
- [48] Han, S., Pool, J., Tran, J., Dally, W.: Learning both weights and connections for efficient neural network. Advances in neural information processing systems **28** (2015)
- [49] Molchanov, P., Tyree, S., Karras, T., Aila, T., Kautz, J.: Pruning convolutional neural networks for resource efficient inference. arXiv preprint arXiv:1611.06440 (2016)
- [50] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
- [51] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
- [52] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986 (2022)
- [53] Li, S., Wang, Z., Liu, Z., Tan, C., Lin, H., Wu, D., Chen, Z., Zheng, J., Li, S.Z.: Moganet: Multi-order gated aggregation network. arXiv preprint arXiv:2211.03295 (2022)



Hongyu Song received the B.E. degree in Software Engineering from Shanxi Agricultural University, Taigu, China, in 2023. He is currently working toward the M.S. degree in Computer Technology with Inner Mongolia University of Technology, Hohhot, China. His research interests include deep learning and image processing.



Bao Shi received the B.E. degree in science from Qinghai University, China, in 2005, and the M.E. and Ph.D. degrees in science from Nagoya City University, Japan, in 2011 and 2015, respectively. He is a professor in the School of Information Engineering, Inner Mongolia University of Technology, China. His research interests include image processing.



Jingping Yang received the bachelor's degree in clinical medicine from Inner Mongolia Medical College in 1986 and the master's degree in respiratory medicine from the Fourth Military Medical University in 1997. Since 2003, he has served as the director of the Department of Respiratory and Critical Care Medicine at Inner Mongolia Baotou Steel Hospital, and concurrently as the director of the Inner Mongolia Institute of Respiratory and Critical Care Medicine. His research and clinical interests include respiratory and critical care medicine.



Xinru Zhang received the B.E. degree in Internet of Things Engineering from Inner Mongolia University of Technology, Hohhot, China, in 2024. She is currently working toward the M.S. degree in Computer Science and Technology with Inner Mongolia University of Technology, Hohhot, China. Her research interests include image processing and computer vision.



Yuan Xue received the B.E. degree in Software Engineering from Langfang Normal University, Langfang, China, in 2024. She is currently working toward the M.S. degree in Computer Technology with Inner Mongolia University of Technology, Hohhot, China. Her research interests include visual recognition, image classification, and computer vision.



Mengfei Wang received the B.E. degree in Computer Science and Technology from Heilongjiang College of Business and Technology, Harbin, China, in 2023. She is currently working toward the M.S. degree in Computer Technology with Inner Mongolia University of Technology, Hohhot, China. Her research interests include deep learning-based image classification and pattern recognition.