PIER M

(Received 21 November 2023, Accepted 3 June 2024, Scheduled 17 July 2024)

A 3-Band Iteration Method to Transfer Knowledge Learned in RGB Pretrained Models to Hyperspectral Domain

Lei Wang[1] and Sailing He[1, 2, *]

[1] *Centre for Optical and Electromagnetic Research, National Engineering Research Center for Optical Instruments, Zhejiang University, Hangzhou 310058, China*
[2] *Taizhou Institute of Zhejiang University, Taizhou, China*

ABSTRACT: We propose a 3-band iteration method to transfer knowledge learned from RGB (red, green, and blue) data pre-trained models to the hyperspectral domain. We demonstrate the classification of a Multi-spectral Choledoch database for cholangiocarcinoma diagnosis. The results show quicker and more stable training progress: over 92% top-1 accuracy in the initial 3 epochs. Some advanced training techniques in the RGB computer vision field can be easily utilized and transferred to the hyperspectral domain without adding more parameters to the original architecture. The computational cost and hardware requirements remain the same. After a majority voting of 3-band combinations, the highest top-1 accuracy on the validation set reached 95.4%, and the highest top-1 accuracy on the test set reached 94.3%. We can directly use our models trained on high-dimensional spectral images to test and infer on RGB color images. We visualized some results by Grad-CAM (Gradient-weighted Class Activation Mapping) on RGB test data, and it shows the transferability of knowledge. We trained the models solely on classification task on spectral data, and these models showed their ability to predict on RGB images with different fields of views. The results indicate good segmentation even when the model has never been trained on any segmentation task.

1. INTRODUCTION

Computer Vision has grown exponentially in the domain of Red-Green-Blue (RGB) imagery. A significant contributor to this vibrant development is the existence of large datasets, for example, ImageNet [1], laion400m [2], laion5b [3] — a dataset consisting of 5.85 billion CLIP [4]-filtered image-text pairs. The success applications of computer vision in the RGB realm primarily lie in the utilization of pre-trained models and transfer learning [5–9]. These advancements rely on large datasets containing hundreds of millions or even billions of labeled images, e.g., 400 million for CLIP [4] and 6.6 billion for BASIC [10]. Hundreds of thousands of models are trained on these datasets with significant computational resources and many training techniques such as knowledge distillation [11], contrastive learning [4], masked image modeling [12], etc., to achieve better performance. A lot of work has been done to update the original released models with better model parameters for higher scores and better performance on benchmarks. The convolutional model we transferred knowledge with is first pre-trained in OpenCLIP [13] on LAION, then fine-tuned on ImageNet-1k to achieve better weights in the model. The transformer-based model we chose learned from a strong CLIP teacher vision encoder [14] via masked image modeling (MIM) pre-training [12] to reconstruct language-aligned vision features, then trained with an intermediate fine-tuning on ImageNet-21K and a subsequent fine-tuning on ImageNet-1K.

Models pre-trained on large amount of data are generally much more robust with generalized knowledge learned and do not tend to overfit or show bias compared to models with smaller training sets. It is more environmentally friendly to transfer the knowledge learned in the model, instead of using more resources to train from scratch [15, 16].

Existing transfer learning techniques for hyperspectral image classification mostly use dimensionality reduction, such as PCA (Principal Component Analysis) or independent component analysis, to match the channels of RGB images. For example, the first three principal components are retained and go to the deep neural network [17]. In this case, the importance of the spectral information is decided by dimensionality reduction projection, not by the deep learning back propagation process during training and optimization. Other works on spectral transfer learning are focused on transfer through different scenes of Remote Sensing [18–20]. Zhang et al. pre-trained a 3-D-CNN for hyperspectral input on 2-D RGB image data sets and then transferred to the target HSI data sets [21], which fails to utilize the pre-trained weights in state-of-the-art RGB models and increases the complexity of neural nets as compared to 3-channel RGB input.

Our method is a universal approach to transfer train RGB pre-trained models to the hyperspectral domain, applicable across any neural network architectures. Convolutional-based models have shown good performance in deep learning computer vision for a long time [22–26], and transformer-based models [12, 14, 27–30] are also rising and proved to even surpass convolutional models' performance in recent years.

* Corresponding author: Sailing He (sailing@zju.edu.cn).

Thus, in the present paper we conducted experiments on both transformer-based and convolutional models using our proposed 3-band iteration method.

Though RGB data and research are the most abundant, multi or hyperspectral imagery involves capturing data at specific frequencies across the electromagnetic spectrum [31–33] and can include both visible light and other bands, such as infrared or ultraviolet. These types of images are commonly used in areas like remote sensing [34], biomedical imaging [35–38], astronomy [39, 40], and agriculture [41–43], offering valuable insights by providing information beyond what human eyes can perceive in standard RGB (Red, Green, Blue) images. However, developing models for multi or hyper-spectral image classification can be challenging for a few reasons. Multi or hyperspectral images are generally much higher-dimensional since they contain multiple layers of information across different spectral bands. This creates computational challenges when processing such images. There is a lack of sufficient labeled datasets in these specific domains due to the complexity and cost associated with capturing and labeling the dataset. Since spectral data can provide more info in its captured spectrum, it is much richer than RGB data. To utilize the knowledge learnt on RGB images in good performance models, we propose a 3-band iteration method to transfer knowledge from RGB pretrained models to the multi or hyper-spectral domain.

In the medical field, multi and hyper-spectral imaging holds immense potential as it provides more detailed information than RGB imagery. Previous methods on processing spectral images often utilize randomly initialized weights in networks like 2D-CNN, 3D-CNN, FC layers, etc., (or refined segmentation annotations [44] with randomly initialized first layer for the mismatch of channel numbers and the channels are handpicked by human decision, and nearly half bands are dropped since high-dimensional hyperspectral images are computationally demanding) and then train the randomly initialized network on the hyper or multi-spectral data. Alternatively, machine learning algorithms [45, 46] are widely employed to process spectral image data pixel by pixel, which can be slower and sometimes require human labeling parts of the image in every image [47] for identification and classification. We evaluated our proposed approach using both transformer-based and convolutional deep neural networks on a multidimensional choledoch dataset to diagnose cholangiocarcinoma. Most prior work focuses on segmentation [44, 48–50] of the dataset using the segmentation annotations provided by class L. However, these annotations are not very accurate [44, 49]. Rough annotations can hurt deep learning models since the models solely learn from the data and are supervised by these annotations. 137 and 145 segment annotations were refined in [44] and [49], respectively, but the refined segmentations are not publicly available. Thus, we applied our method on classification tasks and found that despite no prior exposure to any segmentation annotations, without any training on any segmentation task, our model learned useful transferable knowledge that facilitated tumor areas segmentation solely through classification task training. The direct use of our models trained on spectral images to test and infer on RGB color images demonstrated the transferability of our proposed method. We visualized some Grad-CAM (Gradient-weighted Class Activation Mapping) results on RGB test data to show the knowledge it learned, which has potential to segment by gradient calculations without any segmentation training. This can be meaningful since some medical images are hard to identify and segment out the abnormal areas purely through doctors' eyes before the samples going to histopathologic examinations. It is much easier and quicker for classification labeling than drawing the exact segmentation masks. Datasets built for classification tasks tend to contain more than 100 times labeled data compared to segmentation-annotated datasets. Given this, the flexibility and efficiency of our method could offer benefits in such circumstances. The proposed method is a universal approach to transfer train RGB pre-trained models to the hyperspectral domain, applicable across any neural network architectures. It works universally on transformer-based models and convolutional models, and can be easily adapted to other high-dimensional data.

2. METHODS

We applied our method on a multi-dimensional choledoch dataset containing microscopic hyperspectral images. These images are categorized into three classes: L (samples part of cancer areas), N (samples full of cancer areas), and P (samples without cancer areas). An exemplar hyperspectral image from the dataset is presented in Figure 1.

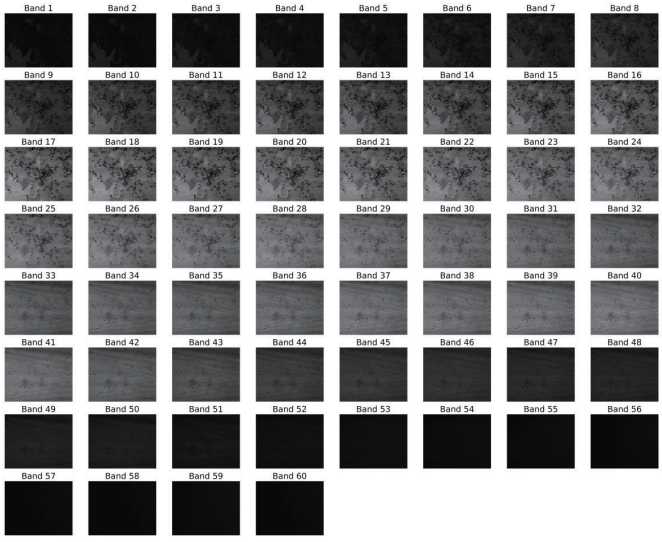

FIGURE 1. The spectral image of choledoch tissue sample 050833_2-20x-roi4 being classified as L (with part cancer areas). 60 spectral bands captured by sCMOS in a wavelength range from 550 nm to 1000 nm with a narrow bandwidth (7.5 nm) via an acousto-optic tunable filter (AOTF).

2.1. 3-Bands Iteration Method

Here we propose a 3-bands iteration method to transfer knowledge in trained models on RGB images. Spectral images are captured with more spectral information as we can see in Figure 2. Though the spectral info can be abundant, the total number of images captured in hyper-spectral cameras are far less

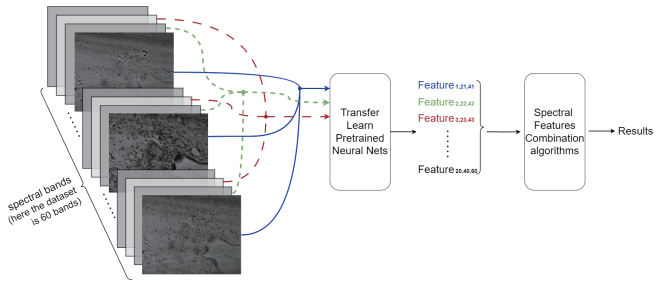

FIGURE 2. The illustration of the transfer training pipelines in our proposed 3-band iteration method for the classification of hyperspectral images.

than RGB color images, which build the strong foundation for the fast development of computer vision in recent years. The spectral bands can be re-grouped as 3-bands per group and fed into the RGB pre-trained models by 3-bands per forward and backward pass to transfer train the neural nets on spectral data tasks. This method works on transformer-based models as well as convolutional models, as we experimented on eva02 models which are based on vision transformers and convnext [26] models that are based on convolutions. After going through the forward pass computation by the transfer learned pre-trained neural nets, we can extract corresponding features from the specific 3-bands fed into the model. The extracted features combination can be fed into the spectral features combination algorithms. The spectral features combination algorithms can be machine learning algorithms, deep learning algorithms, or even both, or can be as simple as removing error-prone features derived from the validation set to achieve a better result on the test set. The spectral features combination algorithm aims to give better results as well as identify important bands for this specific task.

This 3-bands iteration method can also be contributive as a data augmentation method when the 3-bands are more diversely selected and combined. The combination method has $C\,(n, 3)$ number of combinations in total, where n denotes the number of spectral bands.

Take the choledoch hyperspectral dataset for cholangiocarcinoma diagnosis as an example. We split the spectral bands into 20 groups, and each group has 3 bands. Here we choose the 3-bands combination interval to be 20 bands to make the info we feed into the neural net more diverse. Also, we did not feed the 3-bands combination with repeated info, although it could be helpful for data augmentation. Here, to save a lot of computations and make the training process much faster, there is no duplicate band info fed into the neural net within the same epoch. As we can see in Figure 2, in this example dataset, we feed bands $(j,\ j+20,\ j+40)$ iteratively with j ranges from 1 to 20 to the RGB pretrained neural nets.

When the spectral bands iteratively go into the RGB pretrained neural nets, the neural nets are transfer trained on spectral bands through forward and backward pass. After the training process, their corresponding forward features are extracted, and then we can collect the corresponding spectral features and feed them into spectral features combination algorithms to get a result. The spectral features combination algorithms can be as simple as a majority vote and has the potential to make spectral features selection in a data driven and task specific way.

2.2. Dataset

We employed our method on 872 microscopic hyperspectral images of choledoch tissues including 650 images part of cancer areas, 155 images without cancer areas and 67 images full of cancer areas. In this microscopic hyperspectral dataset, the dataset is randomly split into stratified 10 folds, preserving the percentage of samples for each class. We use 80% of the dataset for training, 10% of the dataset for validation and 10% as the test dataset.

2.3. Training

In the training stage, since we are transferring knowledge in the proposed method, we can easily utilize training techniques developed from RGB computer vision field, for instance, data augmentation techniques such as CutMix [51], Mixup [52], AutoAugment [53], RandAugment [54], and Random Erasing [55]. Employing mixed strategy of data augmentation gives us a 2%–4% improvement in top-1 accuracy compared with training processes with no data augmentation.

We trained the model with hyperparameters and optimized them with Bayesian search [56]. The validation performance of algorithms can be seen as a function $f : X \to \mathbb{R}$ of their hyperparameters $\mathbf{x} \in X$. The hyperparameter configuration space X can include both continuous and discrete dimensions. The hyperparameter optimization problem is then defined as finding $\mathbf{x}_\star \in \text{argmax}_{\mathbf{x}\in X} f(\mathbf{x})$.

Due to the intrinsic randomness of most deep learning algorithms, $f(\mathbf{x})$ cannot be observed directly but rather only through noisy observations $y(\mathbf{x}) = f(\mathbf{x}) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma^2_{noise})$. In each iteration i, Bayesian optimization (BO) uses a probabilistic model $p(f|D)$ to model the objective function f based on the already observed data points $D = \{(\mathbf{x}_0, y_0), \ldots, (\mathbf{x}_{i-1}, y_{i-1})\}$. BO uses an acquisition function $a : X \to \mathbb{R}$ based on the current model $p(f|D)$ that trades off exploration and exploitation. Based on the model and the acquisition function, it iterates the following three steps: (1) select the point that maximizes the acquisition function $\mathbf{x}_{new} = \text{argmax}_{\mathbf{x}\in X} a(\mathbf{x})$, (2) evaluate the objective function $y_{new} = f(\mathbf{x}_{new}) + \epsilon$, and (3) augment the data $D \leftarrow D \cup (\mathbf{x}_{new}, y_{new})$ and refit the model. A common acquisition function is the expected improvement (EI) over the currently best observed value $\alpha = \min\{y_0, \ldots, y_n\}$:

$$a(\mathbf{x}) = \int \max(0, \alpha - f(x)) dp(f \mid D). \tag{1}$$

The Tree Parzen Estimator (TPE) [57] is a Bayesian optimization method that uses a kernel density estimator to model the densities

$$\begin{aligned} l(\mathbf{x}) &= p(y < \alpha \| \mathbf{x}, D) \\ g(\mathbf{x}) &= p(y > \alpha \| \mathbf{x}, D) \end{aligned} \tag{2}$$

over the input configuration space instead of modeling the objective function f directly by $p(f|D)$. To select a new candidate $\mathbf{x}_{new}$ to evaluate, it maximizes the ratio $l(\mathbf{x})/g(\mathbf{x})$; Bergstra et al. showed that this is equivalent to maximizing EI in Equation (1). Due to the nature of kernel density estimators,

TPE easily supports mixed continuous and discrete spaces, and model construction scales linearly in the number of data points.

After training, we visualized some results by Grad-CAM (Gradient-weighted Class Activation Mapping) method [58].

3. EXPERIMENTS & RESULTS

In this section, we illustrate the training processes for the classification of microscopic hyperspectral images of cholangiocarcinoma as an example. Figure 3 shows the training process of the model that achieves the highest prediction accuracy on the test set. This model gives a 93.85% top-1 accuracy when evaluated on the test set.

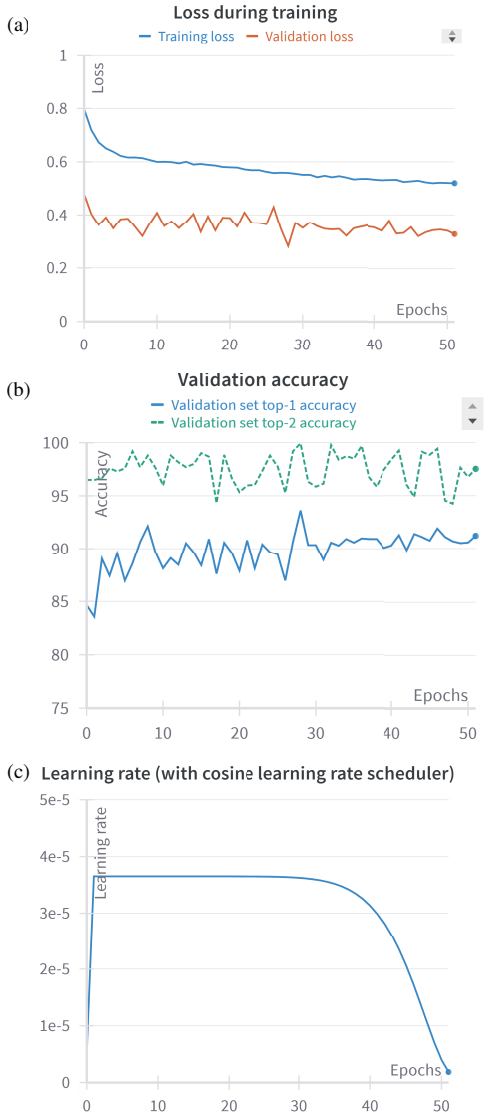

FIGURE 3. The training process of the model that gives a 93.85% top-1 accuracy when evaluated on the test set. (a) The calculated loss during training, here we use the cross-entropy loss as our loss function for training the model. The validation loss is smaller than training loss because of regularizations such as dropout and drop paths. (b) The validation set accuracy when evaluated the model on the validation set in every training epoch. (c) The learning rate during training with a cosine learning rate scheduler to help decay the learning rate for better training.

The hyperparameters for training this good model is obtained by a hyperparameter optimization method: BOHB (Bayesian Optimization and Hyperband). Optimizing hyperparameters can improve a lot on model performance meanwhile maintaining neural nets architectures that are proven to be the state-of-the-art and robust in many benchmarks.

Figure 4 shows one example of the training processes with the hyperparameters obtained through a Bayesian hyperparameter search method. We evaluated models' performance on the validation set after each training epoch. After all models training finished, we tested their performance on the test set. The model with the highest 3-band accuracy on the validation set reached 94.36% top-1 accuracy and this model is transfer-learned from a convnext model. The model with the highest 3-band accuracy on the test set reached 93.85% top-1 accuracy and the test-set-winning model is transfer-learned from an eva02 base model, which is transformer-based. The results of the top-1 accuracy are listed in Table 1. The first 3 columns are the results from our method proposed in this paper. The last column gives the comparison with the previous work [17] where PCA was used to retain the first three principal components to match the three input channels in RGB pre-trained models. From the result, our method gives much better prediction accuracy than the method proposed in [17].

In our transfer method, all the training processes give over 85.11% top-1 accuracy in the initial 3 epochs of training, and the highest top-1 accuracy reached 92.816% in its 2nd epoch of training when validated on the validation set.

Figure 5 shows one of the search space examples for the optimization process of the hyperparameters. Figure 6 shows one of the hyperparameters combinations that gives a low validation loss when the trained model is evaluated on the validation set. Figure 7 shows the importance of the calculated parameters with respect to the validation loss. Figures 4–7 can serve as a starting point to better design the search spaces and further optimize the hyperparameters to achieve better prediction models.

After transfer learning the pretrained RGB neural nets, we can extract features of different bands combinations and design a spectral feature combination algorithm to further process the spectral features and get a result. The algorithm can be very simple such as a majority vote. After the 3-bands combinations majority voting, the highest top-1 accuracy on the validation set reached 95.4%, and the highest top-1 accuracy on the test set reached 94.3%.

Segmentation of tumor area in histopathology is critical. Accurate delineation of tumor margins helps surgical planning and ensures complete tumor resection while minimizing damage to surrounding healthy tissue. Segmentation also facilitates the assessment of tumor heterogeneity, which influences treatment response and the emergence of drug resistance. Using histopathological features derived from segmented tumor areas, clinicians can identify subgroups likely to benefit from targeted therapies, immunotherapy, or adjuvant treatments, and thus optimize therapeutic outcomes and minimize adverse effects. Although the segmentation annotations from this public dataset are not considered very accurate by other work, we visu-

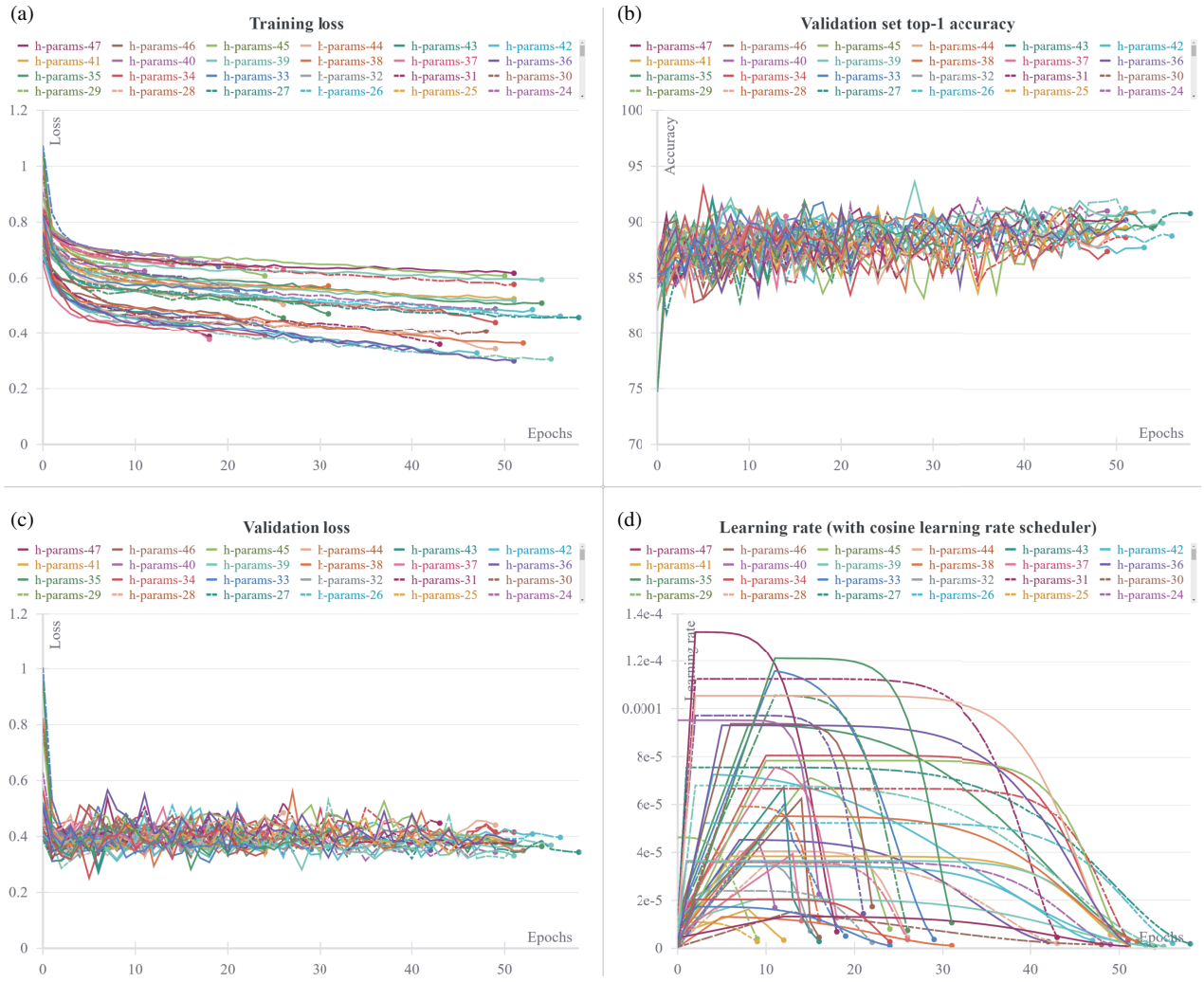

FIGURE 4. One example of the training processes with the hyperparameters obtained through a Bayesian hyperparameter optimization method. (a) shows every training loss calculated for each epoch; (b) shows the top-1 accuracy when the trained model is evaluated on the validation set for each training epoch; (c) shows the evaluated loss when the trained model is evaluated on the validation set for each training epoch; (d) shows the scheduled learning rate as epochs grow.

TABLE 1. Highest prediction accuracy from our transfer trained model and comparison with previous work [17].

	3-band accuracy	transfer trained model	Majority voting accuracy	PCA + pretrained model [17]
Validation set	94.36%	Convnext (Convolution-based)	95.4%	90.70%
Test set	93.85%	Eva02 (transformer-based)	94.3%	88.764%

alized some Grad-CAM (Gradient-weighted Class Activation Mapping) results on RGB test data in Figure 8. From Figure 8 we can see that the learned knowledge is inter-transferable between the spectral images and the RGB images. Also, the direct use of our models trained on spectral images to test and infer on RGB color images demonstrated the transferability of our proposed method. We have trained our model only on spectral images, and now we can test our model on some RGB color images with different field of views, since RGB color images are easier and cheaper to collect as compared with hyper or multi-spectral data.

When we calculate the Gradient-weighted Class Activation Mapping (Grad-cam), shown in Figure 8, for our models trained on the spectral data, we can see that they also perform well on RGB data that it has never seen during the training. The model has learned useful and effective information: it is only trained on the classification task (but not directly on the RGB data) and yet from the grad-cam visualization results we can see that it even gives good segmentations of tumor areas without any segmentation annotations training. Since the results we showed here are derived from naïve grad-cam calculations, extracted only from a single layer and without any small value filters and

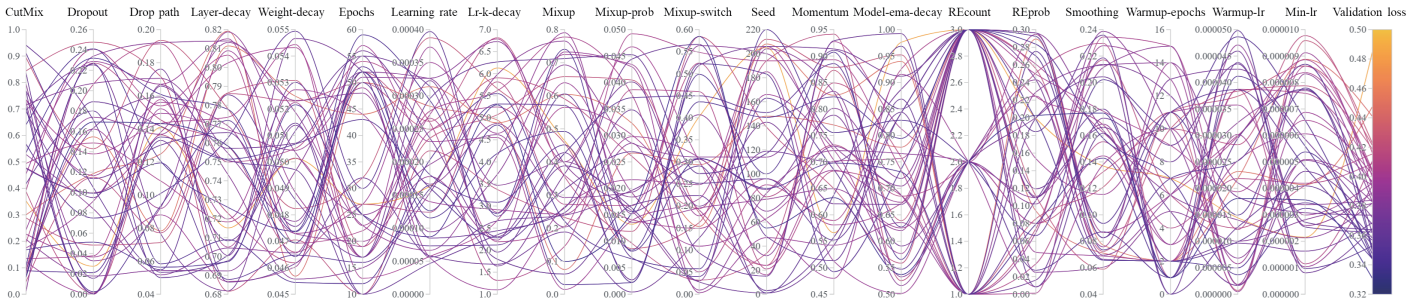

FIGURE 5. One example of the search spaces of the Bayesian hyperparameters optimization process. The left 20 columns are the hyperparameters settings for training the models. Each hyperparameter was set a range for the Bayesian optimization search to sample from and start a training process with this hyperparameters combination. The last column is the corresponding validation loss, with the combination of hyperparameters connected in a line. Our hyperparameters for optimization including: CutMix, dropout, drop path, layer decay, weight decay, epochs, learning rate, learning rate k-decay, mixup, mixup probability, mixup switch probability, random seed, momentum, model-ema-decay, random erase count, random erase probability, smoothing, warmup epochs, warmup leaning rate, minimum learning rate (from left to right), to achieve better the prediction model.

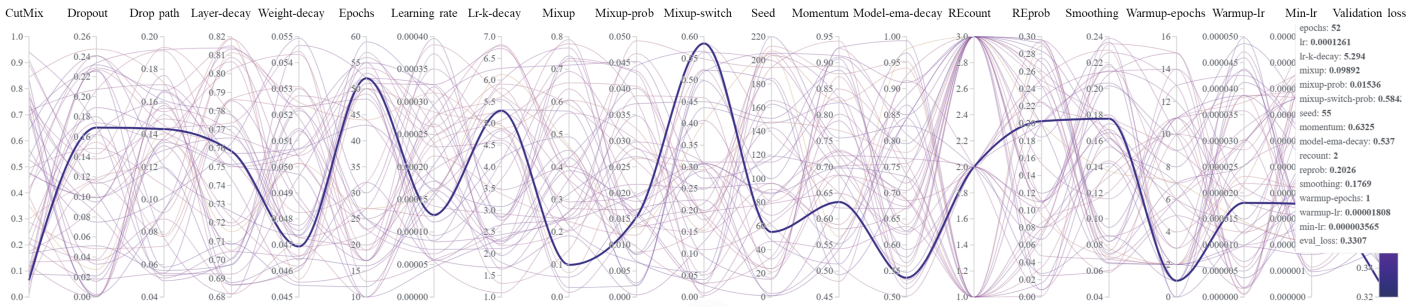

FIGURE 6. One combination of the hyperparameters that gives a low validation loss when the trained model is evaluated on the validation set.

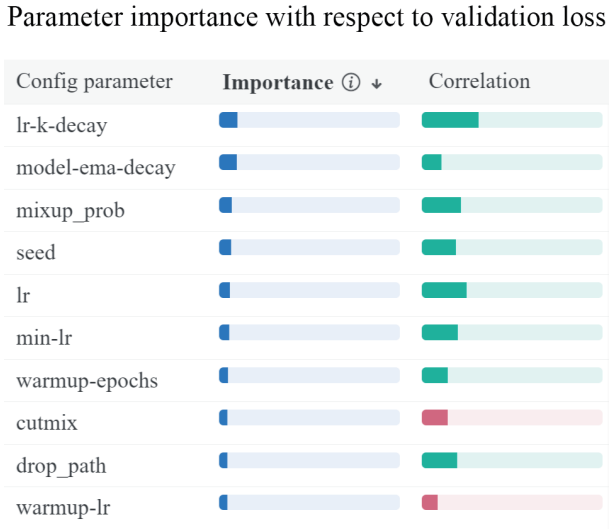

FIGURE 7. The importance and correlation of the calculated parameters with respect to the evaluation loss. The importance metric is derived by training a random forest with the hyperparameters as inputs, the metric as the target output and report the feature importance values for the random forest. This can help fine-tune the hyperparameter searches by showing which of the hyperparameters matter the most in terms of predicting model performance.

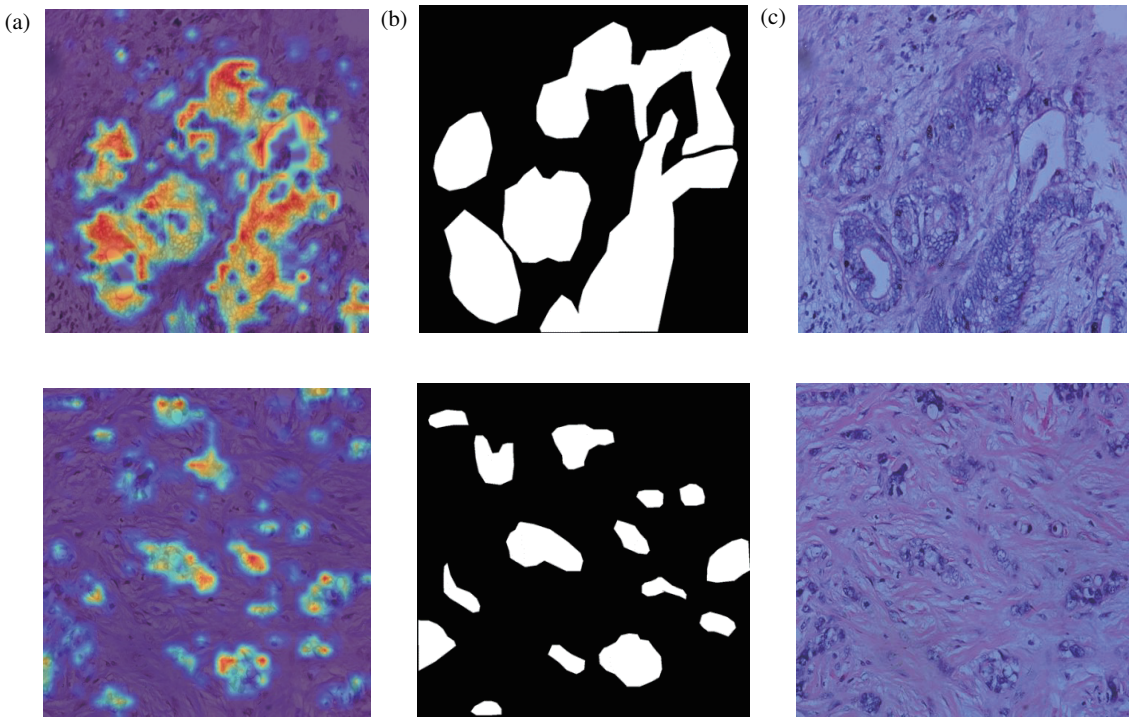

FIGURE 8. (a) Our Grad-cam results for the model only trained on spectral images and tested on RGB color images with different fields of views to show the learned knowledge is inter-transferable. (b) The original segmentation annotation provided in the database (not the refined version). (c) The tested RGB color images fed into our model. From the results we found that models only trained on classification task, can make good indications or even segment properly without any training on segmentation annotations. This finding can be meaningful for medical data which is easy to identify its class by histopathology and biochemical examinations, but is hard to see and identify the exact segmentation annotations without histopathological examinations.

no smoothing at all, it has great potential to be improved in future works.

4. DISCUSSIONS AND CONCLUSION

From the experiments and results, we can see the effectiveness of the proposed 3-bands iteration method to transfer knowledge learned on large amount of RGB data pretrained models to the spectral domain. The demonstrated experiments and results showed the benefits of this transfer training pipeline.

Since our approach works on transformer-based models as well as CNNs, and can be easily adapt to other high-dimensional data transfer learning process, it can be adjusted and applied to many other fields.

The spectral features combination algorithm in our proposed method has great potential to design more sophisticated features combination algorithms and give a task-specific data-driven method to infer back on the importance of raw features. This module of spectral features combination algorithms can be explored more in future works.

We also demonstrated good transferability of our method by training on high-dimensional spectral images and inferring on RGB color test images directly. The Grad-CAM results that we present indicate the potential to segment tumor areas solely through classification task training without any segmentation annotations. This can be meaningful since some medical images are hard to identify. Also, it is much easier and quicker for classification labeling than drawing the exact segmentation masks. Datasets built for classification tasks tend to contain more than 100 times labeled data compared to segmentation-annotated datasets. Given this, the flexibility and efficiency of our method could offer many benefits and be a worthy direction for the future work.

ACKNOWLEDGEMENT

This work is partially supported by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (Nos. 2022C03051, 2023C03083 and 2023C03135), the National Key Research and Development Program of China (Nos. 2022YFC2010000 and 2022YFC3601000) and the National Natural Science Foundation of China (No. 11621101). The authors gratefully thank Dr. Julian Evans of Zhejiang University for helpful discussion.

REFERENCES

[1] Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255, Miami, FL, USA, 2009.

[2] Schuhmann, C., R. Vencu, R. Beaumont, R. Kaczmarczyk, C. Mullis, A. Katta, T. Coombes, J. Jitsev, and A. Komatsuzaki, "LAION-400M: Open dataset of CLIP-filtered 400 million image-text pairs," *ArXiv Preprint ArXiv:2111.02114*, 2021.

[3] Schuhmann, C., R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, P. Schramowski, S. Kundurthy, K. Crowson, L. Schmidt, R. Kaczmarczyk, and J. Jitsev, "LAION-5B: An open large-scale dataset for training next generation image-text models," *Advances in Neural Information Processing Systems*, Vol. 35, 25 278–25 294, 2022.

[4] Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning, PMLR*, 8748–8763, 2021.

[5] West, J., D. Ventura, and S. Warnick, "Spring research presentation: A theoretical foundation for inductive transfer," *Brigham Young University, College of Physical and Mathematical Sciences*, Vol. 1, No. 08, 2007.

[6] George Karimpanal, T. and R. Bouffanais, "Self-organizing maps for storage and transfer of knowledge in reinforcement learning," *Adaptive Behavior*, Vol. 27, No. 2, 111–126, 2019.

[7] Maitra, D. S., U. Bhattacharya, and S. K. Parui, "CNN based common approach to handwritten character recognition of multiple scripts," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 1021–1025, Tunis, Tunisia, 2015.

[8] Bird, J. J., J. Kobylarz, D. R. Faria, A. Ekárt, and E. P. Ribeiro, "Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG," *IEEE Access*, Vol. 8, 54 789–54 801, 2020.

[9] Kabir, H. M. D., M. Abdar, A. Khosravi, S. M. J. Jalali, A. F. Atiya, S. Nahavandi, and D. Srinivasan, "Spinalnet: Deep neural network with gradual input," *IEEE Transactions on Artificial Intelligence*, Vol. 4, No. 5, 1165–1177, 2023.

[10] Pham, H., Z. Dai, G. Ghiasi, K. Kawaguchi, H. Liu, A. W. Yu, J. Yu, Y.-T. Chen, M.-T. Luong, Y. Wu, M. Tan, and Q. V. Le, "Combined scaling for zero-shot transfer learning," *Neurocomputing*, Vol. 555, 126658, 2023.

[11] Hinton, G., O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *ArXiv Preprint ArXiv:1503.02531*, 2015.

[12] Bao, H., L. Dong, S. Piao, and F. Wei, "Beit: Bert pre-training of image transformers," *ArXiv Preprint ArXiv:2106.08254*, 2021.

[13] Ilharco, G., M. Wortsman, R. Wightman, C. Gordon, N. Carlini, R. Taori, A. Dave, V. Shankar, H. Namkoong, J. Miller, *et al.*, "Openclip," 2021.

[14] Fang, Y., W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang, and Y. Cao, "Eva: Exploring the limits of masked visual representation learning at scale," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19 358–19 369, 2023.

[15] Lannelongue, L., J. Grealey, and M. Inouye, "Green algorithms: Quantifying the carbon footprint of computation," *Advanced science*, Vol. 8, No. 12, 2100707, 2021.

[16] Tamburrini, G., "The AI carbon footprint and responsibilities of AI scientists," *Philosophies*, Vol. 7, No. 1, 4, 2022.

[17] Zhu, S., J. Zhang, M. Chao, X. Xu, P. Song, J. Zhang, and Z. Huang, "A rapid and highly efficient method for the identification of soybean seed varieties: Hyperspectral images combined with transfer learning," *Molecules*, Vol. 25, No. 1, 152, 2020.

[18] Zhong, C., J. Zhang, S. Wu, and Y. Zhang, "Cross-scene deep transfer learning with spectral feature adaptation for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, 2861–2873, 2020.

[19] Yang, B., S. Hu, Q. Guo, and D. Hong, "Multisource domain transfer learning based on spectral projections for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, 3730–3739, 2022.

[20] Khaki, S., H. Pham, and L. Wang, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning," *Scientific Reports*, Vol. 11, No. 1, 11132, 2021.

[21] Zhang, H., Y. Li, Y. Jiang, P. Wang, Q. Shen, and C. Shen, "Hyperspectral classification based on lightweight 3-D-CNN with transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 8, 5813–5828, 2019.

[22] Krizhevsky, A., I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, Vol. 25, 2012.

[23] Simonyan, K. and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ArXiv Preprint ArXiv:1409.1556*, 2014.

[24] He, K., X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, 2016.

[25] Tan, M. and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning, PMLR*, 6105–6114, 2019.

[26] Liu, Z., H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR*, 2022.

[27] Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," *ArXiv Preprint ArXiv:2010.11929*, 2020.

[28] Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10 012–10 022, 2021.

[29] Tu, Z., H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. M. Li, "Multi-axis vision transformer," in *Proceedings of the Computer Vision — ECCV 2022: 17th European Conference*, 459–479, 2022.

[30] Fang, Y., Q. Sun, X. Wang, T. Huang, X. Wang, and Y. Cao, "Eva-02: A visual representation for neon genesis," *ArXiv Preprint ArXiv:2303.11331*, 2023.

[31] Chang, C.-I., *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Springer Science & Business Media, 2003.

[32] Grahn, H. and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*, John Wiley & Sons, 2007.

[33] Hagen, N. A. and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Optical Engineering*, Vol. 52, No. 9, 090901, 2013.

[34] Sumbul, G., M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *IGARSS 2019 — 2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904, Yokohama, Japan, 2019.

[35] Lu, G. and B. Fei, "Medical hyperspectral imaging: A review," *Journal of Biomedical Optics*, Vol. 19, No. 1, 010901, 2014.

[36] Zhang, Q., Q. Li, G. Yu, L. Sun, M. Zhou, and J. Chu, "A multidimensional choledoch database and benchmarks for cholan-

giocarcinoma diagnosis," *IEEE Access*, Vol. 7, 149 414–149 421, 2019.

[37] Wang, Q., S. Fernandes, G. O. S. Williams, N. Finlayson, A. R. Akram, K. Dhaliwal, J. R. Hopgood, and M. Vallejo, "Deep learning-assisted co-registration of full-spectral autofluorescence lifetime microscopic images with h&e-stained histology images," *Communications Biology*, Vol. 5, No. 1, 1119, 2022.

[38] Jiao, C., Z. Lin, Y. Xu, and S. He, "Noninvasive Raman imaging for monitoring mitochondrial redox state in septic rats," *Progress In Electromagnetics Research*, Vol. 175, 149–157, 2022.

[39] McLean, I. S., *Beating the Atmosphere*, 39–75, 2008.

[40] Majaess, D., "Discovering protostars and their host clusters via WISE," *Astrophysics and Space Science*, Vol. 344, 175–186, 2013.

[41] Lacar, F. M., M. M. Lewis, and I. T. Grierson, "Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia," in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, 2875–2877, Sydney, NSW, Australia, 2001.

[42] Vermeulen, P., P. Flémal, O. Pigeon, P. Dardenne, J. A. F. Pierna, and V. Baeten, "Assessment of pesticide coating on cereal seeds by near infrared hyperspectral imaging," *Journal of Spectral Imaging*, Vol. 6, No. a1, 1–7, 2017.

[43] Zea, M., A. Souza, Y. Yang, L. Lee, K. Nemali, and L. Hoagland, "Leveraging high-throughput hyperspectral imaging technology to detect cadmium stress in two leafy green crops and accelerate soil remediation efforts," *Environmental Pollution*, Vol. 292, 118405, 2022.

[44] Zhou, Z., S. Qiu, Y. Wang, M. Zhou, X. Chen, M. Hu, Q. Li, and Y. Lu, "Swin-spectral transformer for cholangiocarcinoma hyperspectral image segmentation," in *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–6, Shanghai, China, 2021.

[45] Zhu, H., J. Luo, J. Liao, and S. He, "High-accuracy rapid identification and classification of mixed bacteria using hyperspectral transmission microscopic imaging and machine learning," *Progress In Electromagnetics Research*, Vol. 178, 49–62, 2023.

[46] Xu, Z., Y. Jiang, J. Ji, E. Forsberg, Y. Li, and S. He, "Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning," *Optics Express*, Vol. 28, No. 21, 30 686–30 700, 2020.

[47] Wang, M., Y. Xu, Z. Wang, and C. Xing, "Deep margin cosine autoencoder based medical hyperspectral image classification for tumor diagnosis," *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, 2023.

[48] Shan, X., T. Ma, A. Gu, H. Cai, and Y. Wen, "TCRNet: Make transformer, CNN and RNN complement each other," in *ICASSP 2022 — 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1441–1445, Singapore, Singapore, 2022.

[49] Dai, K., Z. Zhou, S. Qiu, Y. Wang, M. Zhou, M. Li, and Q. Li, "A generative data augmentation trained by low-quality annotations for cholangiocarcinoma hyperspectral image segmentation," in *2023 International Joint Conference on Neural Networks (IJCNN)*, 01–09, Gold Coast, Australia, 2023.

[50] Zhan, G., Y. Uwamoto, and Y.-W. Chen, "Hyperunet for medical hyperspectral image segmentation on a choledochal database," in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, 1–5, IEEE, Las Vegas, NV, USA, 2022.

[51] Yun, S., D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 6023–6032, Seoul, Korea, 2019.

[52] Zhang, H., M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," *ArXiv Preprint ArXiv:1710.09412*, 2017.

[53] Cubuk, E. D., B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation strategies from data," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 113–123, Long Beach, CA, USA, 2019.

[54] Cubuk, E. D., B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 702–703, Seattle, WA, USA, 2020.

[55] Zhong, Z., L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, 13 001–13 008, 2020.

[56] Falkner, S., A. Klein, and F. Hutter, "BOHB: Robust and efficient hyperparameter optimization at scale," in *Proceedings of the 35th International Conference on Machine Learning, PMLR*, 1437–1446, 2018.

[57] Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, 2546–2554, 2011.

[58] Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626, 2017.