# Detection of Basal Cell Carcinoma in Whole Slide Images

Hongyan Xu[1,2], Dadong Wang[2(✉)], Arcot Sowmya[1], and Ian Katz[3]

[1] School of Computer Science and Engineering, University of New South Wales, Kensington, Australia
{hongyan.xu,a.sowmya}@unsw.edu.au
[2] Data61, The Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
Dadong.wang@csiro.au
[3] Southern Sun Pathology Pty Ltd, Thornleigh, Australia
ian.katz@southernsun.com.au

**Abstract.** Basal cell carcinoma (BCC) is a prevalent and increasingly diagnosed form of skin cancer that can benefit from automated whole slide image (WSI) analysis. However, traditional methods that utilize popular network structures designed for natural images, such as the ImageNet dataset, may result in reduced accuracy due to the significant differences between natural and pathology images. In this paper, we analyze skin cancer images using the optimal network obtained by neural architecture search (NAS) on the skin cancer dataset. Compared with traditional methods, our network is more applicable to the task of skin cancer detection. Furthermore, unlike traditional unilaterally augmented (UA) methods, the proposed supernet Skin-Cancer net (SC-net) considers the fairness of training and alleviates the effects of evaluation bias. We use the SC-net to fairly treat all the architectures in the search space and leveraged evolutionary search to obtain the optimal architecture for a skin cancer dataset. Our experiments involve 277,000 patches split from 194 slides. Under the same FLOPs budget (4.1G), our searched ResNet50 model achieves 96.2% accuracy and 96.5% area under the ROC curve (AUC), which are 4.8% and 4.7% higher than those with the baseline settings, respectively.

**Keywords:** Basal cell carcinoma (BCC) · whole-slide pathological images · deep learning · neural architecture search (NAS)

## 1 Introduction

Skin cancer, the most prevalent cancer globally, has seen increasing incidences over recent decades [1]. It constitutes a third of all cancer diagnoses, affecting one in five Americans [2]. Basal cell carcinoma (BCC), comprising 70% of cases, has surged by 20–80% in the last 30 years, exerting a significant healthcare strain.


H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 263–272, 2023.
https://doi.org/10.1007/978-3-031-43987-2_26

Timely BCC diagnosis is crucial to avoid complex treatments. Although histological evaluation remains the gold standard for detection [3], deep learning and computer vision advancements can streamline this process. Scanned traditional histology slides result in whole slide images (WSIs) that can be analyzed by deep learning models, significantly easing the histological evaluation burden. Recent advancements underscore the promise of this approach [4–6].

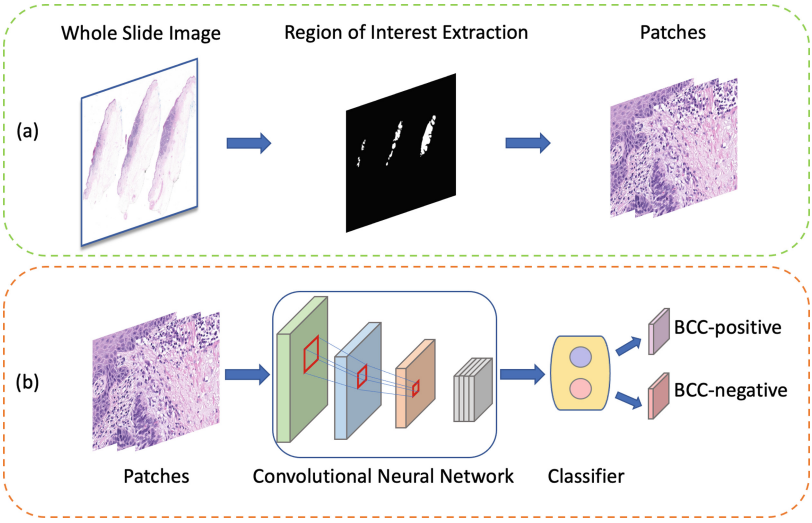


**Fig. 1.** Overall framework of the proposed model. (a) Region of interest (ROI) extraction and patch generation, and (b) patch detection and WSI classification.

Existing skin cancer detection methods [7–9] typically employs models like Inception Net and ResNet, designed for natural images like those in the ImageNet dataset. The significant variance in pathology and natural images can compromise these models' accuracy. Neural architecture search (NAS) addresses this issue by auto-designing superior models [10–13], exploring a vast architecture space. However, current NAS methods often overlook fairness in architecture ranking, impeding the discovery of top-performing models.

In this study, we utilized the NAS approach to identify the optimal network for skin cancer detection. To improve the efficiency and accuracy of the search, we developed a new framework named SC-net, which focuses on identifying highly valuable architectures. We observed that conventional NAS methods often overlook fairness ranking during the search, hindering the search for optimal solutions. Our SC-Net framework addresses this by ensuring fair training and precise ranking. The efficacy of SC-net was confirmed by our experimental results, with our ResNet50 achieving 96.2% top-1 accuracy and 96.5% AUC, outperforming baseline methods by 4.8% and 4.7% respectively.

Figure 1 shows the proposed framework, which integrates two modules. Module (a) extracts the region of interest (ROI) from WSI and generates patches,

while Module (b) uses optimal model architecture from NAS to analyze features from patches and generate classifications.

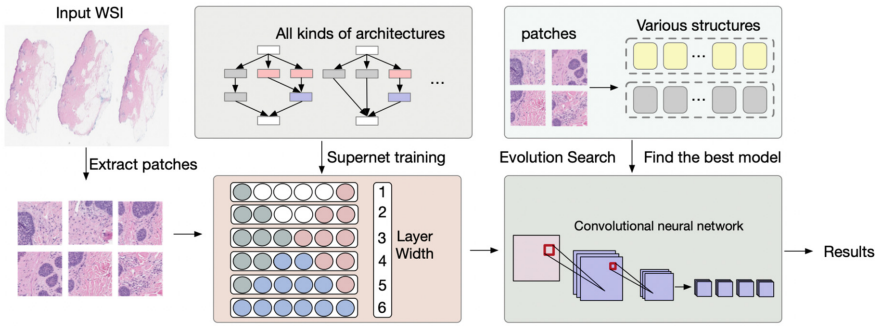


**Fig. 2.** The schematic diagram of the proposed method. Module (a) extracts ROI from WSI, creating $224 \times 224$ patches. Module (b) uses these patches to train and search optimal structure within a supernet via an evolutionary algorithm, yielding dataset predictions.

## 2 Methods

The proposed method (Fig. 2) involves dividing the input WSI into patches for training a supernet and the search for optimal architectures [14]. Section 2.2 provides further details about the supernet. A balanced evolutionary algorithm is then used to select the optimal structure from the search space, with the candidate structures' performance evaluated using mini-batch patch data. We evaluate the searched architectures on the skin cancer dataset.

### 2.1 One-Shot Channel Number Search

To extract an optimal architecture $\gamma \in \mathcal{G}$ from a vast search space $\mathcal{G}$, a weight-sharing strategy [15–17] is used to prevent training from scratch. The search leverages a supernet $\mathcal{S}$ with weights $\mathcal{W}$, with each path $\gamma$ inheriting weights from $\mathcal{W}$. This makes one-shot NAS a two-step optimization process: supernet training and architecture search. The original dataset is typically split into training $\mathcal{D}_t$ and validation datasets $\mathcal{D}v$. The weights $\mathcal{W}$ of the supernet $\mathcal{S}$ are trained by uniformly sampling the network width $d$ and optimizing the sub-network with weights $wd \subset \mathcal{W}$. The optimization function is defined as follows:

$$\mathcal{W}^* = \arg\min_{w_d \in \mathcal{W}} \mathop{\mathbb{E}}_{d \in U(\mathcal{D})} [\![\mathcal{L}_t(w_d; \mathcal{S}, d, \mathcal{D}_t]\!]) \tag{1}$$

where $U(\mathcal{D})$ is a uniform distribution of network widths, $\mathbb{E}$ is the expected value of random variables, and $\mathcal{L}_t$ is the training loss function. Then, the optimal

network width $d^*$ corresponds to the network width with the best performance (*e.g.* classification accuracy) on the validation dataset, *i.e.*,

$$d^* = \arg\max_{d \in \mathcal{D}} Acc(d, w_d^*; \mathcal{W}^*, \mathcal{S}, \mathcal{D}_v), \text{s.t. } FLOPs(d) \leq F_p, \tag{2}$$

where $F_p$ is the resource budget of FLOPs. The search for Eq. 2 can be efficiently performed by various algorithms, such as random or evolutionary search [18]. Afterward, the performance of the searched optimal width $d^*$ is analyzed by training from scratch.

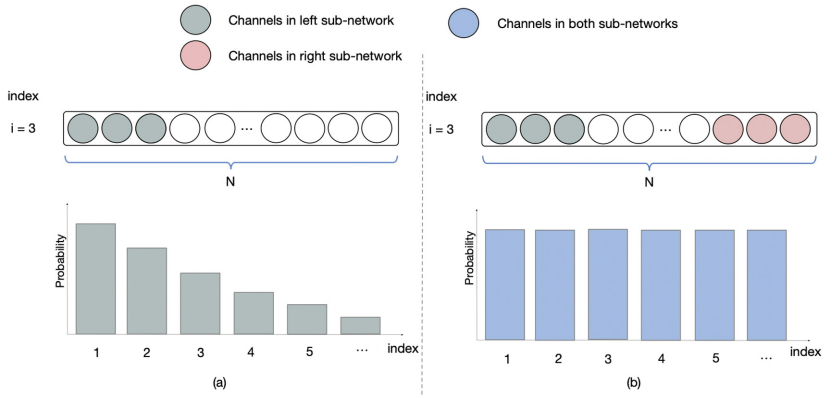


**Fig. 3.** Schematic diagram comparing (a) the UA method and (b) our method. In the UA principle, some channels are trained twice while others are trained only once or not at all, leading to channel training unfairness and evaluation bias. In contrast, our proposed method ensures that all channels are trained evenly (twice) by training both the width $d$ and its complementary width.

### 2.2 SC-Net as a Balanced Supernet

Current approaches for neural architecture search [19–21] often employ a unilaterally augmented (UA) principle to evaluate each width, resulting in unfair training of channels in the supernet. As illustrated in Fig. 3(a), to search for a dimension $d$ at a layer with a maximum of $n$ channels, the UA principle assigns the left $d$ channels in the supernet to indicate the corresponding architecture as

$$\gamma_A(d) = [1 : d], d \leq n \tag{3}$$

where $\gamma_A(d)$ means the selected $d$ channels from the left (smaller-index) side.

However, the UA principle leads to channel training imbalance in the supernet due to its constraints, as illustrated in Fig. 3(a). Channels with smaller indices are used for various sizes, resulting in over-training of the left channel kernels since widths are uniformly sampled. The unfairness can be quantified

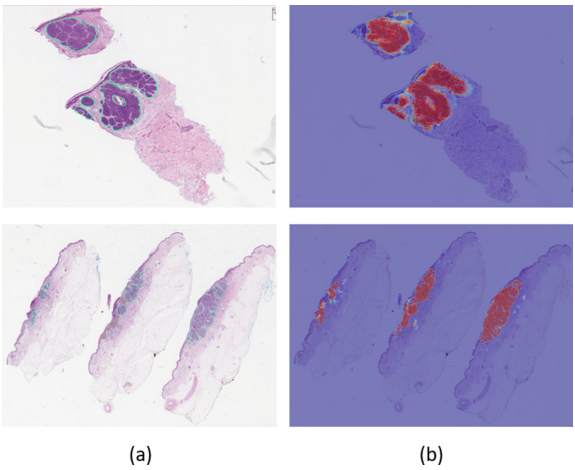


**Fig. 4.** Example of prediction probability heatmaps produced by the s_ResNet50 model. Column (a) shows the original WSIs that contain annotations represented by green circles. Column (b) displays the prediction probability heatmaps of our ResNet50 model. (Color figure online)

by $\mathcal{T}$, which represents how often a channel is utilized, reflecting its level of training. Given a layer has a maximum of $n$ channels, the $\mathcal{T}$ for the $i$-th channel under the UA principle is

$$\mathcal{T}(i) = n - i + 1 \tag{4}$$

Correspondingly, the probability of $i$ -th channel being trained can be expressed as $P_i = \frac{n-i+1}{n}$. Therefore, channels closer to the left will get more attempts during training, which leads the degree of training to vary widely between channels. This introduces evaluation bias and leads to sub-optimal results.

To mitigate evaluation bias on width, we propose a new SC-net that promotes the fairness of channels during training. As shown in Fig. 3(b), in the proposed supernet, each width is simultaneously evaluated by the sub-networks corresponding to the left and right channels. This can be seen as two identical networks $S_l$ and $S_r$ that are bilaterally coupled and evaluated using the UA principle, but counting channels in reverse order. Therefore, the number of all channels used for evaluating the width $d$ can be expressed as:

$$\mathcal{H}(d) = \mathcal{H}^l_{UA}(d) \uplus \mathcal{H}^r_{UA}(d) \tag{5}$$

$$= [1 : d] \uplus [(n - d + 1) : (n - d)], \tag{6}$$

where $\uplus$ represents the union of two lists with repeatable elements. In detail, the left channel in $S_l$ follows the same UA principle setup as in Eq. (3), while for the right channel in $S_r$, we count channels from the right $\mathcal{H}^r(d) = [(n - d + 1) :$

$(n-d)]$. Therefore, the training degree of each channel is the sum of the two supernets $S_l$ and $S_r$. Since the channels are counted from the right within $S_r$, the training degree of the $d$-th channel on the left corresponds to the training degree of the $(n-d+1)$-th channel on the right in Eq. (4). Therefore, the training degree $\mathcal{T}(d)$ of the $d$-th channel in our proposed method is

$$\mathcal{T}(d) = \mathcal{T}_{UA}(d) + \mathcal{T}_{UA}(n+1-d) \tag{7}$$

$$= (n-d+1) + (n+1-n-1+d) = n+1 \tag{8}$$

Therefore, the training degree $\mathcal{T}$ for each channel will always be equal to the same constant value of the width, independent of the channel index, ensuring fairness in terms of channel (filter) levels. Thus the network width can be fairly ranked using our network.

### 2.3 Balanced Evolutionary Search with SC-Net

Using a trained SC-net, the architecture can be evaluated. However, the search space involved in NAS is large, with more than $10^{20}$ possible architectures, requiring an evolutionary search using the multi-objective NSGA-II algorithm to improve the search performance. During the evolutionary search, the width $d$ of each network is represented by the average precision of its corresponding left and right paths in the supernet $\mathcal{S}$, as shown in Eq. (9). The optimal width (not subnetwork) is determined as the one that achieves the best performance when trained from scratch. Here, $\mathcal{S}_l$ and $\mathcal{S}_r$ refer to the two paths of $\mathcal{S}$ that correspond to the width $d$ during the training process.

$$Acc(\mathcal{W}, d, \mathcal{D}_v) = \frac{1}{2}\left(Acc(\mathcal{S}_l, d; \mathcal{D}_v) + Acc(\mathcal{S}_r, d; \mathcal{D}_v)\right) \tag{9}$$

## 3 Experiments

### 3.1 Experiment Settings

**Table 1.** Generated dataset split

| | BCC-positive | | BCC-negative | |
|---|---|---|---|---|
| Training | WSI | Patch | WSI | Patch |
| | 118 | 132,981 | 37 | 90,291 |
| Testing | WSI | Patch | WSI | Patch |
| | 30 | 31,651 | 9 | 22,838 |

The dataset, comprised of 194 skin slides acquired from the Southern Sun Pathology laboratory, includes 148 BCC cases and 46 other types (common nevus,

SCC), all manually annotated by a dermatopathologist. BCC slides served as positive samples and the rest as negatives. These slides were scanned at ×20 magnification with a 0.44 µm pixel size using a Leica Aperio AT2 Scanner. The patient data were separated between training and testing to prevent overlap. Details are shown in Table 1. The experimental setup involved training models on two NVIDIA RTX A6000 GPUs using PyTorch. These models, initialized from a zero-mean Gaussian with standard deviation $\sigma = 0.001$, were trained for 200 epochs with a batch size of 256. Training used the Adam optimizer with a dynamic learning rate reduction strategy, starting with a learning rate of 5e-5 following a cosine schedule.

**Table 2.** Performance comparison on skin cancer dataset

| Type | Model | FLOPs | Parameters | Acc | Se | Sp | $F_1$ | AUC |
|---|---|---|---|---|---|---|---|---|
| WSI analysis-Related | Tian *et al.* [22] | 4.1G | 25.5M | 92.5 | 92.3 | 92.1 | 92.2 | 92.7 |
| | Hekler *et al.* [9] | 4.1G | 25.5M | 93.4 | 92.9 | 92.2 | 92.8 | 93.5 |
| | Jiang *et al.* [23] | 2.9G | 27.2M | 92.1 | 91.2 | 91.7 | 91.4 | 92.4 |
| NAS-Related | MetaPruning [18] | 4.1G | 25.5M | 94.1 | 93.2 | 92.7 | 93.3 | 94.2 |
| | AutoSlim [24] | 4.1G | 25.5M | 93.8 | 92.9 | 92.2 | 93.0 | 93.0 |
| Ours | ori_ResNet50 | 4.1G | 25.5M | 91.4 | 90.2 | 90.4 | 90.5 | 91.8 |
| | s_ResNet50 | 4.1G | 27.2M | **96.2** | **94.7** | **95.8** | **95.2** | **96.5** |
| | ori_MobileNetV2 | 300M | 3.5M | 86.7 | 86.2 | 85.8 | 86.2 | 87.3 |
| | s_MobileNetV2 | 300M | 4.0M | 91.9 | 90.4 | 91.7 | 90.7 | 92.4 |

### 3.2 Performance Evaluation

We validated our algorithm using the curated skin cancer dataset and SC-net as a supernet, testing both heavy and light models. We performed a search on ResNet50 and MobileNetV2 models, compared against original ResNet50 (ori_ResNet50) and MobileNetV2 (ori_MobileNetV2) models as baselines. The resulting models are denoted as s_ResNet50 and s_MobileNetV2.

**Comparison with Related Methods.** To ensure a fair comparison on our dataset, we selected several papers in the field of pathological image analysis, such as [9,22,23], as well as others using the UA principle, such as [18,24].

**Evaluation Metrics.** Our model was evaluated on: (1) Accuracy (Acc): percentage of correct classifications. (2) Sensitivity (Se): proportion of true positives identified. (3) Specificity (Sp): proportion of true negatives identified. (4) F1 Score (F1): precision and recall's harmonic mean, indicating label alignment. (5) AUC: ROC curve area, reflecting the false/true positive rate trade-off.

As shown in Table 2, the s_ResNet50 model outperformed in all metrics, showing 4.8%, 4.5%, 5.4%, 4.7% and 4.7% improvements in accuracy, sensitivity, specificity, $F_1$ Score, and AUC, respectively, over ori_ResNet50, and surpassing

**Table 3.** Performance of searched models with different searching methods.

| Evaluator | Searching | | Models | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SC-net | Greedy search | Evolutionary search | ResNet50 | | | | MobileNetV2 | | | |
| | | | Acc | Se | Sp | AUC | Acc | Se | Sp | AUC |
| | ✓ | | 93.8 | 92.9 | 92.2 | 93.0 | 89.7 | 90.3 | 89.9 | 90.2 |
| ✓ | ✓ | | 94.9 | 93.7 | 93.1 | 95.3 | 90.7 | 89.8 | 90.0 | 91.2 |
| | | ✓ | 94.0 | 93.5 | 92.7 | 94.2 | 89.8 | 90.2 | 90.5 | 90.5 |
| ✓ | | ✓ | **96.2** | **94.7** | **95.8** | **96.5** | 91.9 | 90.4 | 91.7 | 92.4 |

the Hekler *et al.* [9] method by 2.8%, 1.8%, 3.6%, 2.4%, and 3.0%. It also compared favorably to the s_MobileNetV2, adding only 1.7M and 0.5M parameters, respectively. In terms of UA principle method, s_ResNet50 advanced by 2.1%, 1.5%, 3.1%, 1.9%, and 2.3% in the same metrics compared to MetaPruning [18], showcasing our method's efficacy while keeping model complexity reasonable.

**Visualization of Probability Heatmaps.** Figure 4 demonstrates an example of the prediction probability heatmaps produced by the s_ResNet50 model. A comparison of the labeled areas in Column (a) and the red areas in Column (b) indicates that the predicted areas are generally similar in scope to the corresponding labeled areas. This suggests that the model is accurately identifying regions of interest for BCC diagnosis.

### 3.3 Ablation Study

**Effect of SC-net as a Supernet.** Table 3 presents our experiments testing the SC-net on ResNet50 and MobileNetV2 using various supernets and search methods. The SC-net (second row) under greedy search improved accuracy by 1.1% (ResNet50) and 1.0% (MobileNetV2), while with evolutionary search, accuracy increased by 1.8% (ResNet50) and 2.1% (MobileNetV2). These results highlight SC-net's effectiveness as a supernet in bolstering evaluation and search performance.

**Generalization Ability of SC-net.** We tested our model's generalization on the ChestMNIST and DermaMNIST subsets of MedMNISTv2 [25], following established protocols. As shown in Table 4, our s_ResNet50 surpassed the original ResNet50 on all datasets, gaining 2.3% and 1.8% more AUC on ChestMNIST and DermaMNIST respectively, proving the model's robust generalization.

**Table 4.** Performance comparison on ChestMNIST and DermaMNIST datasets

| Model | ChestMNIST | | DermaMNIST | |
|---|---|---|---|---|
| | AUC | ACC | AUC | ACC |
| ResNet18 | 76.8 | 94.7 | 91.7 | 73.5 |
| ResNet50 | 76.9 | 94.7 | 91.3 | 73.5 |
| auto-sklearn | 64.9 | 77.9 | 90.2 | 71.9 |
| AutoKeras | 74.2 | 93.7 | 91.5 | 74.9 |
| Google AutoML | 77.8 | 94.8 | 91.4 | 76.8 |
| s_ResNet50 | **79.2** | **95.5** | **93.1** | **77.8** |

## 4 Conclusion and Future Work

In this paper, we introduce SC-net, a novel NAS framework for skin cancer detection in pathology images. By formulating SC-net as a balanced supernet, we ensure fair ranking and treatment of all potential architectures. With SC-net and evolutionary search, we obtained optimal architectures, achieving 96.2% Top-1 and 96.5% accuracy on a skin cancer dataset, improvements of 4.8% and 4.7% over baselines. Future work will apply our approach to larger datasets for wider-scale validation.

## 5 Compliance with Ethical Standards

This study was performed in line with the principles of the Declaration of Helsinki. Ethics approval was granted by CSIRO Health and Medical Human Research Ethics Committee (CHMHREC). The ethics approval number is 2021_030_LR, and the validity period is from 07 Apr 2021 to 31 Dec 2024.

## References

1. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012. JAMA Dermatol. **151**(10), 1081–1086 (2015)
2. Stern, R.S.: Prevalence of a history of skin cancer in 2007: results of an incidence-based model. Arch. Dermatol. **146**(3), 279–282 (2010)
3. Xu, G., et al.: CAMEL: a weakly supervised learning framework for histopathology image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10682–10691 (2019)
4. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. **5**(6), 555–570 (2021)
5. Sharma, Y., Shrivastava, A., Ehsan, L., Moskaluk, C.A., Syed, S., Brown, D.: Cluster-to-conquer: a framework for end-to-end multi-instance learning for whole slide image classification. In: Medical Imaging with Deep Learning, pp. 682–698. PMLR (2021)

6. Xiang, T., et al.: DSNet: a dual-stream framework for weakly-supervised gigapixel pathology image analysis. IEEE Trans. Med. Imaging **41**(8), 2180–2190 (2022)
7. Lu, C., Mandal, M.: Automated analysis and diagnosis of skin melanoma on whole slide histopathological images. Pattern Recogn. **48**(8), 2738–2750 (2015)
8. Van Zon, M., et al.: Segmentation and classification of melanoma and nevus in whole slide images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 263–266. IEEE (2020)
9. Hekler, A., et al.: Pathologist-level classification of histopathological melanoma images with deep neural networks. Eur. J. Cancer **115**, 79–83 (2019)
10. Su, X., You, S., Wang, F., Qian, C., Zhang, C., Xu, C.: BCNet: searching for network width with bilaterally coupled network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2175–2184 (2021)
11. Liu, C., et al.: Progressive neural architecture search. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 19–34 (2018)
12. Xie, B., et al.: Multi-scale fusion with matching attention model: a novel decoding network cooperated with NAS for real-time semantic segmentation. IEEE Trans. Intell. Transp. Syst. **23**(8), 12622–12632 (2021)
13. Su, X., et al.: Vision transformer architecture search, arXiv preprint arXiv:2106.13700 (2021)
14. Cha, S., Kim, T., Lee, H., Yun, S.-Y.: Supernet in neural architecture search: a taxonomic survey, arXiv preprint arXiv:2204.03916 (2022)
15. Liu, H., Simonyan, K., Yang, Y.: DARTS: differentiable architecture search, arXiv preprint arXiv:1806.09055 (2018)
16. Xie, L., et al.: Weight-sharing neural architecture search: a battle to shrink the optimization gap. ACM Comput. Surv. (CSUR) **54**(9), 1–37 (2021)
17. Chen, X., Xie, L., Wu, J., Wei, L., Xu, Y., Tian, Q.: Fitting the search space of weight-sharing NAS with graph convolutional networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 8, pp. 7064–7072 (2021)
18. Liu, Z., et al.: Metapruning: meta learning for automatic neural network channel pruning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3296–3305 (2019)
19. Chen, M., Peng, H., Fu, J., Ling, H.: Autoformer: searching transformers for visual recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12270–12280 (2021)
20. Wan, A., et al.: Fbnetv2: differentiable neural architecture search for spatial and channel dimensions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12965–12974 (2020)
21. Yan, Z., Dai, X., Zhang, P., Tian, Y., Wu, B., Feiszli, M.: FP-NAS: fast probabilistic neural architecture search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15139–15148 (2021)
22. Tian, Y., et al.: Computer-aided detection of squamous carcinoma of the cervix in whole slide images, arXiv preprint arXiv:1905.10959 (2019)
23. Jiang, Y., et al.: Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network. Br. J. Dermatol. **182**(3), 754–762 (2020)
24. Yu, J., Huang, T.: Autoslim: towards one-shot architecture search for channel numbers, arXiv preprint arXiv:1903.11728 (2019)
25. Yang, J., et al.: MedMNIST v2-a large-scale lightweight benchmark for 2D and 3D biomedical image classification. Sci. Data **10**(1), 41 (2023)