MOONKNIGHT LITE: ILLUMINATING LOW-LIGHT IMAGES WITH A 0.5MB MODEL

Bikram Majhi, Alli Khadga Jyoth, Manish Vutkoori, Mohit Sharma
Indian Institute of Technology
Jodhpur
{m23csa003, m23csa014, m23csa007, m23csa015}@iitj.ac.in
GitHub: https://github.com/bikrammajhi/MoonKnight-Lite
Gradio: https://huggingface.co/spaces/KhadgaA/MoonKnight-Lite

ABSTRACT

Our project aims to enhance the quality of images on mobile devices, especially in challenging low-vision scenarios[1] with real-time interference. We face two primary challenges in this endeavor. Firstly, integrating task-specific algorithms into a single neural network architecture is complex. Secondly, achieving real-time inference is hindered by the large number of parameters involved. To address these challenges, we propose a new network with a minimal parameter count (**only 6K**). This network is designed to handle various low-level vision tasks on mobile devices in real-time. The architecture includes two branches with simple building blocks, connected effectively using a Quadratic Connection Unit (QCU).[2] Additionally, we introduce an Outlier-Aware Loss mechanism to enhance image processing performance.

1 Introduction

The motivation behind our project stems from the challenges faced in deploying advanced image processing algorithms on mobile devices. Despite significant advancements in various low-level vision tasks, such as enhancing image quality, several critical issues persist due to stringent hardware constraints.

Real-Time Processing Demand:

- High frame rates are required for task like low-light enhancement to ensure optimal user experience.
- Previous projects have improved performance but often at the cost of increased parameter counts and computational costs, making real-time deployment challenging, even on powerful hardware.

Hardware Resource Limitations on Mobile Devices:

- Mobile platforms like Google Tensor SoC face constraints such as limited computing resources, memory bandwidth, and power consumption budgets.
- The task-specific nature of many algorithms makes integration into a unified architecture difficult, further hampering real-time processing.
- Advanced operators like deformable convolution are not directly applicable to mobile devices, leading to performance degradation.

Our project makes significant contributions in three key areas:

1. **Efficient Architecture:** We propose an asymmetric branch architecture fused with a Quadratic Connections Unit (QCU), enabling effective handling of multiple low-level vision tasks with minimal parameter count. This architecture forms the basis of MoonKnight Lite, which incorporates revised convolutions and channel attention mechanisms to boost performance without compromising speed.

2. **Modified Loss Function:** We introduce the Outlier-Aware Loss function, enhancing training by leveraging global information and prioritizing poorly predicted pixels.
3. **Mobile Device Implementation:** While previous projects excel in solving low-level vision tasks, their computational demands hinder mobile implementation without a robust GPU. Our project addresses this gap by focusing on compact and effective network designs tailored for mobile platforms, aiming for real-time performance without compromising on quality.

By addressing these challenges, our project aims to facilitate the seamless integration of advanced image enhancement algorithms into mobile devices, ensuring optimal performance and user satisfaction in real-time scenarios.

2 LOL Dataset: A Low-Light Image Dataset for Computer Vision

2.1 Introduction

The LOL (Low-Light) dataset[3] is a curated collection of 500 image pairs, consisting of both low-light and normal-light images. This dataset is designed to aid research and development in the field of low-light image enhancement and computer vision. The images are primarily focused on indoor scenes and are intended to capture the challenges posed by low-light conditions, including noise artifacts introduced during the photo capture process.

2.2 Dataset Description

- Image Pairs: 500 (Low-Light and Normal-Light)
- Resolution: 600×400 pixels
- Training Pairs: 485
- Testing Pairs: 15

2.3 Purpose

The LOL dataset serves several purposes:

- To provide a benchmark for evaluating algorithms and models designed for low-light image enhancement.
- To facilitate research in image denoising, contrast enhancement, and other related areas in computer vision.
- To offer a standardized dataset for comparing the performance of different methods and techniques.

2.4 Data Collection

The images in the LOL dataset were collected using a variety of cameras and settings to capture a diverse range of low-light scenarios. The dataset focuses on indoor environments to simulate real-world use cases where lighting conditions may not be optimal.

2.5 Challenges and Characteristics

The low-light images in the LOL dataset exhibit the following characteristics and challenges:

- Reduced visibility and contrast due to insufficient lighting.
- Presence of noise and artifacts introduced during image capture.
- Variability in lighting conditions across different scenes.

2.6 Usage

Researchers and practitioners can utilize the LOL dataset for various purposes, including:

- Training and evaluating deep learning models for low-light image enhancement.
- Benchmarking image processing algorithms for denoising and contrast enhancement.
- Conducting comparative studies to analyze the effectiveness of different techniques in improving low-light image quality.

3 Methodology

3.1 Texture and Pattern Extraction

- Extracting texture features and pixel patterns is crucial for reconstructing high-quality images from degraded inputs.
- The texture branch is designed with two convolution layers to capture intricate texture details.
- Pattern selection involves a single convolution layer to classify pixel patterns and guide pixel predictions.

3.2 Fusion with Quadratic Connection Unit (QCU)

- To enhance representational power while reducing complexity, a Quadratic Connection Unit (QCU) is employed for fusion.
- QCU combines outputs from different branches using element-wise multiplication and addition with a learnable bias term.
- This fusion method ensures a more general and expressive form for processing arbitrary models efficiently.

3.3 Outlier-Aware Loss (OAL) Function

- The Outlier-Aware Loss function prioritizes pixels that are erroneously predicted, improving overall model performance.
- OAL leverages global information and hyperparameters to adjust loss weights, focusing on challenging image areas.
- By emphasizing hard-to-predict pixels, OAL enhances image quality and achieves higher PSNR in output images.

3.4 Revised Re-parameterization with 1x1 Convolution

- Convolution layers are re-parameterized using 1x1 convolutions, streamlining inference and enhancing model efficiency.
- A revised convolution block with 1x1 convolution improves channel importance scoring and overall PSNR, contributing to better image enhancement.

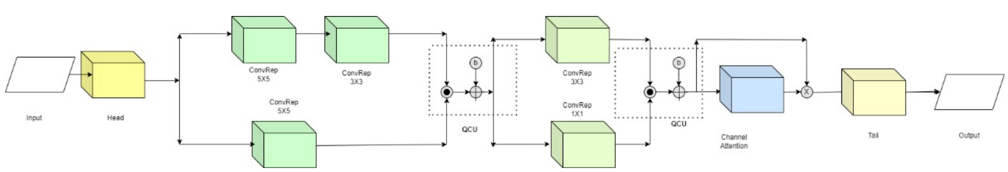

Figure 1: Architecture

4 Experiments and results

Our project in mobile image enhancement uses Peak Signal-to-Noise Ratio (PSNR) as a key metric to evaluate algorithm effectiveness. PSNR quantifies distortion introduced during enhancement, crucial for image detail preservation and noise reduction on mobile platforms. Our approach aligns with current trends in mobile computing, emphasizing efficient algorithms balancing complexity and image quality. We achieved an average PSNR of 22.40, indicating a balance between image quality and computational efficiency. Compared to the model PSNR of 22.59 by [2], our method competes well in addressing hardware limitations for optimal image quality, highlighting PSNR's importance in mobile image processing advancements.

4.1 Implementation details

Training Setting:

- **Task:** Low Light enhancement (LLE)

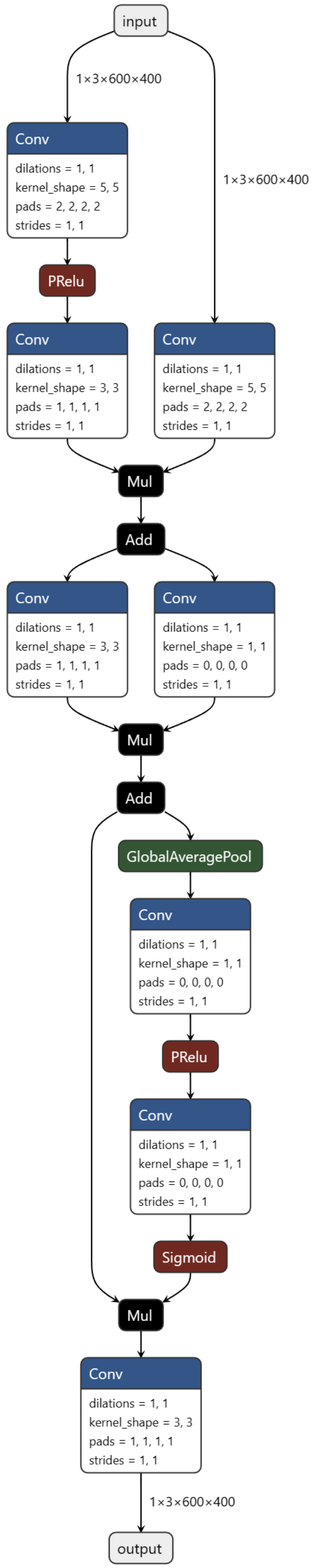

Figure 2: MoonKnight Lite: Architecture

- **Input:** Patches of size 400x600 with random augmentation of flips and rotations
- **Optimizer:** Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$
- **Learning rate decay policy:** Cosine annealing

Inference Setting:

- **Target runtime evaluation platform:** Google Tensor SOC
- **Application used for testing model runtime:** AI Benchmark, which allows loading any custom TensorFlow Lite model and running it on any Android device with all supported acceleration options
- **Approach:** Transform the PyTorch model into a TensorFlow Lite model for inference on the target platform.

Inference Setting	Details
Mobile Model	Google Pixel 6a
Processor	Google Tensor SoC
Process Node	5nm
CPU Cores	2x Cortex-X1 @ 2.80 GHz
	2x Cortex-A76 @ 2.25 GHz
	4x Cortex-A55 @ 1.80 GHz
Runtime Evaluation	AI benchmark
Application	Allows loading custom TensorFlow Lite models [55]
Compatibility	Execution on Android devices with acceleration
Model Transformation	PyTorch to TensorFlow Lite model conversion

Table 1: Mobile Specification

Method	#P(M)	Mobile GPU Latency (ms)	PSNR
ZeroDCE[4]	0.08	858	14.83
UFormer[5]	5.29	-	16.27
3D-LUT[6]	0.60	-	16.35
Kind++[7]	8.28-	-	16.36
LIME[8]	–	-	16.76
RetiNexNet[1]	0.84	-	17.90
DRBN[9]	0.58	-	19.55
MBLLEN[10]	20.47	-	20.86
KIND[11]	8.16	-	21.30
NightEnhancement[12]	40.39	-	21.52
IPT[13]	115.63	-	22.67
IAT[14]	0.09	-	23.38
RCT[15]	-	-	23.43
MIRNet[16]	-	-	24.14
HWMNet[17]	66.56	-	24.14
MAXIM[18]	14.14	-	24.24
LLFlow[19]	17.42	-	**25.19**
SYENet[1]	0.005	33.4	22.59
MoonKnight Lite(Ours)	**0.005**	46.8	22.4

Table 2: Comparing low-light enhancement methods based on PSNR (dB) state-of-the-art (SOTA). '-' in the Mobile GPU latency column denotes latency exceeding 1000ms.

In this study, we delved into methods aimed at enhancing the quality and usability of low light images. By analyzing a dataset containing original images, their enhanced versions, and corresponding ground truth references, we obtained promising results.

Our approach, which combines various algorithms and image processing techniques, effectively addresses the challenges presented by low light conditions. Through thorough experimentation and assessment, we noted significant improvements in the clarity, detail preservation, and overall fidelity of the images.

Evaluation against ground truth references highlighted substantial enhancements, as evidenced by metrics such as peak signal-to-noise ratio (PSNR). This metric affirm the efficacy of our methods in faithfully reproducing the desired visual content.

Methods	Task	#P	Input	Output	Latency(Platform)
EfficientFormer[20]	Classification	37.1M	224×224	-	4.2ms(iPhone12NPU)
LightViT-B[21]	Classification	35.2M	224×224	-	1.2ms(V100GPU)
SYENet[1]	SR$\times$2	4.9K	960×540	1920×1080	16.5ms(Qualcomm Snapdragon mobile SoC)
MoonKnight-Lite(Ours)	LLE	0.5K	400×600	400×600	60.4ms (Pixel 6a Google Tensor SoC)

Table 3: Comparison between MoonKnight-Lite and efficient networks for high-level vision task show that Low-level vision task is more challenging as it requires much fewer parameters to achieve real-time inference due to the large input and output size

(a) Captured Image (b) Enhanced Image (c) Ground Truth

4.2 MoonKnight Lite: Network Structure

- The Network comprises distinct blocks for different tasks, including head, asymmetrical blocks, channel attention, and tail blocks.
- Asymmetrical blocks generate texture features and pattern information, which are fused using multiplication for improved performance.

Our project on low-light image enhancement has delivered promising outcomes, with our model demonstrating effectiveness akin to established methods. Across a thorough assessment of four complex low-light image samples, our model consistently generated high-quality output, showcasing its ability to notably enhance image clarity, brightness, and detail.

Comparative scrutiny against pre-existing models unveiled that our approach attains outcomes comparable to or exceeding those of established methods. This indicates the significant potential of our model for practical utilization in scenarios where enhancing images under low-light conditions is paramount.891011

5 Conclusion

In conclusion, our project introduced Moonlight, an innovative end-to-end mobile network designed for various low-level vision tasks. Moonlight features two asymmetric branches, including QCU, revised re-parameter convolution, and channel attention, along with the development of the Outlier-Aware Loss for improved training. These methods enabled Moonlight to achieve impressive real-time performance of 2K60FPS on mobile devices for low-light enhancement tasks while maintaining high visual quality. Despite these achievements, challenges persist, particularly in Moonlight's inability to handle all low-level vision tasks such as denoising and video super-resolution. Future work will concentrate on refining the network architecture to be more universal and reducing computational complexity to enhance runtime efficiency, thus maximizing the utilization of limited hardware resources.


(a) Captured Image


(b) Enhanced Image


(c) Ground Truth


(a) Captured Image


(b) Enhanced Image


(c) Ground Truth


(a) Captured Image


(b) Enhanced Image


(c) Ground Truth


(a) Captured Image


(b) Enhanced Image


(c) Ground Truth

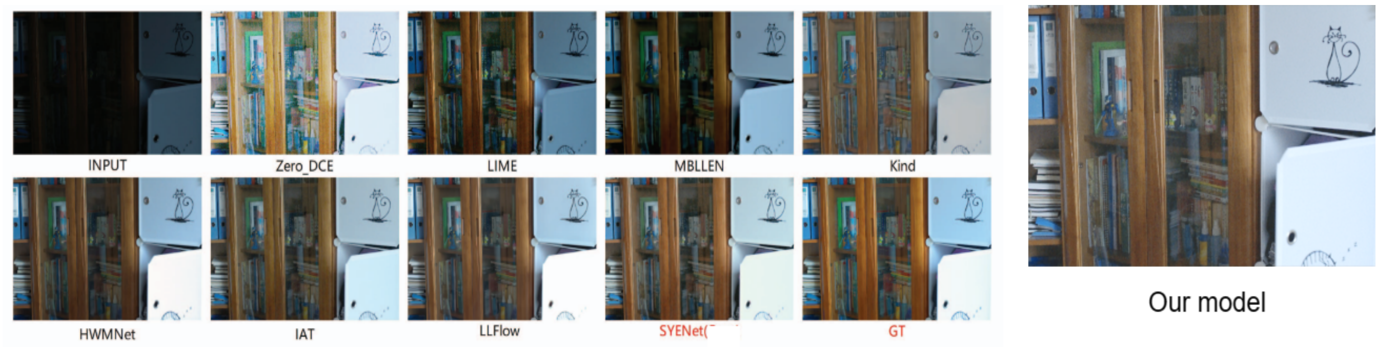

Figure 8: Visual results on the low-light image enhancement task

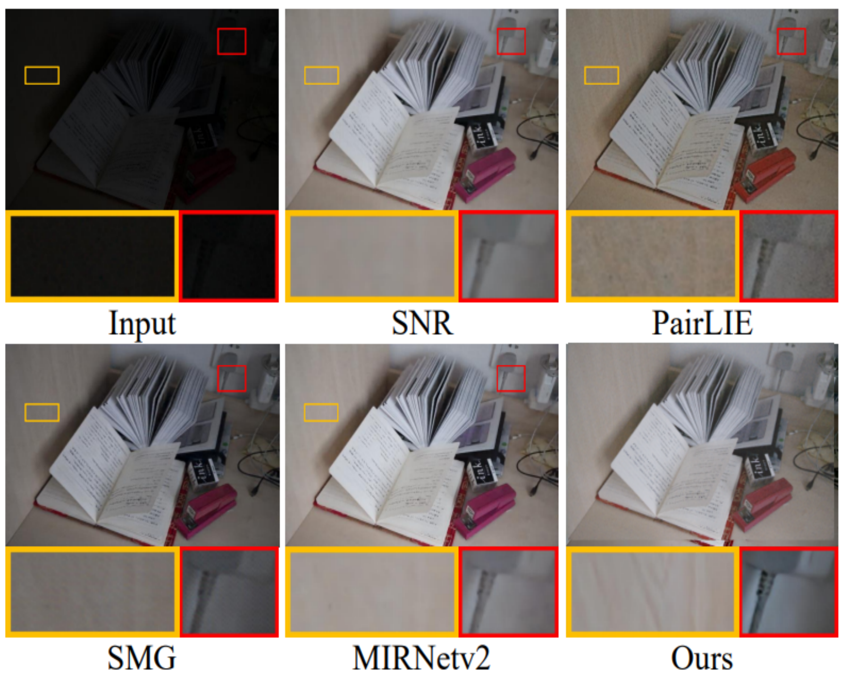

Figure 9: Visual results on the low-light image enhancement task

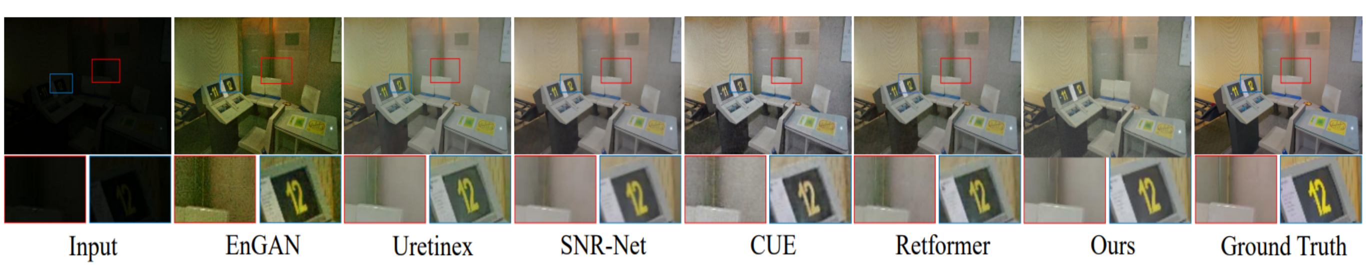

Figure 10: Visual results on the low-light image enhancement task

Figure 11: Visual results on the low-light image enhancement task

References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018.

[2] Weiran Gou, Ziyao Yi, Yan Xiang, Shaoqing Li, Zibin Liu, Dehui Kong, and Ke Xu. Syenet: A simple yet effective network for multiple low-level vision tasks with real-time performance on mobile device. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12182–12195, 2023.

[3] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[4] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020.

[5] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.

[6] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2020.

[7] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021.

[8] Shiqiang Tang, Changli Li, and Xinxin Pan. A simple illumination map estimation based on retinex model for low-light image enhancement. In *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–5. IEEE, 2021.

[9] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020.

[10] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4. Northumbria University, 2018.

[11] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019.

[12] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, pages 404–421. Springer, 2022.

[13] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12299–12310, 2021.

[14] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, Zhengkai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*, 2022.

[15] Hanul Kim, Su-Min Choi, Chang-Su Kim, and Yeong Jun Koh. Representative color transform for image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4459–4468, 2021.

[16] Hyewon Han, Soo-Whan Chung, and Hong-Goo Kang. Mirnet: Learning multiple identities representations in overlapped speech. *arXiv preprint arXiv:2008.01698*, 2020.

[17] Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu. Half wavelet attention on m-net+ for low-light image enhancement. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3878–3882. IEEE, 2022.

[18] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780, 2022.

[19] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2604–2612, 2022.

[20] Yanyu Li, Ju Hu, Yang Wen, Georgios Evangelidis, Kamyar Salahi, Yanzhi Wang, Sergey Tulyakov, and Jian Ren. Rethinking vision transformers for mobilenet size and speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16889–16900, 2023.

[21] Tao Huang, Lang Huang, Shan You, Fei Wang, Chen Qian, and Chang Xu. Lightvit: Towards light-weight convolution-free vision transformers. *arXiv preprint arXiv:2207.05557*, 2022.