Satellite imagery-based crop type segmentation for small fields using deep learning neural networks

Deo Uwimpuhwe
MSECE 2024
duwimpuh@andrew.cmu.edu

Gustave Bwirayesu
MSECE 2024
gbwiraye@andrew.cmu.edu

Eric Maniraguha
MSIT 2024
emanirag@andrew.cmu.edu

Jules Bienvenue Himbaza
MSECE 2024
hjulesbi@andrew.cmu.edu

December 2023

1 Abstract

Agriculture, a foundational pillar of developing economies, plays a crucial role in sub-Saharan Africa, employing a significant workforce and contributing over 15% to the GDP of African nations. Despite progress, a noticeable technology gap hampers the optimal use of advanced tools for strategic planning and crop monitoring in the region. Additionally, limited labeled data availability and small-scale farming systems have resulted in a shortage of research on crop type mapping in Africa, especially on small fields. Our project addresses this gap by harnessing deep learning on remote sensing data, emphasizing the practicality of satellite imagery in resource-constrained regions for precise crop type mapping.

Incorporating attention mechanisms, which have recently gained prominence in vision-based applications, our approach adopts Swin-Unet, a state-of-the-art model for segmentation tasks. Swin-Unet processes satellite images collected in Ghana in 2016 and 2017, extracting features and generating a class mask with spatial resolution akin to the input images. The results showcase a significant improvement, with the Swin-Unet model achieving a validation accuracy of 93.6%. This study pioneers the application of state-of-the-art deep learning algorithms in satellite imagery analysis, highlighting the potential of advanced technologies in advancing precision agriculture in resource-constrained regions.[1]

2 Introduction

Agriculture is the key driver of many economies in the developing world [1]. Agriculture employs most people in sub-Saharan Africa [2] and contributes, on average, to over 15% of the GDP of African countries [3]. The rising population and chronic hunger in Africa are the prime causes of the increasing food demand that calls governments and private sectors to invest in agriculture. Both the Maputo Declaration of 2023 [4] and the Malabo Declaration of 2014 [5] highlight the initiative of African heads of state to urgently invest in agriculture. Today, Africa made a significant improvement in agriculture; but there is a big gap in using state-of-the-art technologies to facilitate strategic planning and crop monitoring.

[1] We all worked collaboratively on abstract

The increase in Earth observation satellites has triggered numerous applications, particularly in agriculture and environmental monitoring. The importance of satellite-based applications is heightened by key factors such as temporal, spatial, and spectral resolutions[6][7]. These three pieces of information, provided by satellites, play a crucial role in understanding land covers, including crops. Temporal resolution refers to the revisit time of a satellite over a specific region of interest. Spatial resolution defines the area on the ground represented by a single pixel, while spectral resolution is pivotal in providing reflectance for different bands. These three characteristics are essential considerations in most applications during the data collection process. This technology contributes to the data availability, allowing enhanced data analysis. The project's essential purpose is to help in precision agriculture by applying deep learning algorithms on remote sensing data of crop types. The Swin-Unet employed in this work, takes the images from 3 different satellites leveraging their unique properties to map crop types.

Satellite images are seen as a feasible source of data for resource-constrained regions or remote areas that are hard to reach. This also addresses the high cost of sensors needed to collect agricultural farm data. Information on crop type in certain areas helps concerned decision-makers in planning for agricultural inputs, thus increasing the yield. This will also help decision-makers to mobilize for agrarian resources.[2][3]

3 Literature Review

Deep learning contributes significantly to the rise of remote sensing-based applications [8] [9] [10]. CNN (Convolutional Neural Network) is a pivotal architecture for many vision applications [11]. Semantic segmentation is one of these applications which classifies each image pixel to create class labels [12] [13].

Without overlooking the fame of CNN-based architectures in image processing, very recently, attention mechanism has gained much importance in computer vision applications [14]. In [15], A. Dosovitskiy et al. built on the transformer concept in Natural Language Processing to establish vision transformer (ViT) architecture which outperformed ResNets architectures on different datasets. Similar to word tokenization in NLP, the image is split into fixed-scale patches that are flattened to create patch embeddings. This poses two shortcomings, it is computationally expensive to learn high-resolution images and it is not applicable for vision tasks like semantic segmentation that require pixel-wise prediction. In response, [**?**] introduces a Swin-transformer that hierarchically extracts features based on shifted windows. This was also extended to U-Net architecture as first introduced as Swin-Unet in [16] where the encoder is Swin transformer while the decoder is also Swin transformer but with patch expanding instead of patch merging.

Crop type mapping using satellite imagery is a challenging task considering the environmental factors such as clouds, shadows, radiometric errors and unavailability of data. To cater for data scarcity, efforts have been made by different organizations to ensure data availability. Satellite missions like EU Copernicus sentinel mission by European Space Agency provides open source satellite based data of different temporal, spectral and spatial resolutions[17]. Nonetheless, the problem remains ground truth to validate and train models that can be built for remote sensing applications.

With deep learning, remote sensing data continues to attract worldwide cutting-edge technology that significantly impacts agriculture production [6]. Multiple researchers leverage these tech-

[2] Deo Uwimpuhwe highlighted the economic significance of agriculture.

[3] Eric Maniraguha emphasized the transformative role of Earth observation satellites.

nological resources in regions where there is huge satellite data and large-scale farming systems. This, therefore, supports building high-performing deep learning models. In Idaho, One of the largest USA states, Ehsan Raei et al. [18] applied U-Net architecture on high spatial-resolution (1m) remote sensing images to segment four irrigation systems. Their model with the Resnet-34 backbone, achieved a state-of-the-art performance of 72% to 82% for the validation dataset. In China Baodi County, Zhenrong Du et al [19] used the DeepLab3+ model to semantically segment crop area based on high-resolution images of the WorldView-2 (WV-2) satellite. The model presented an overall accuracy of close to 95% and comparatively performs better than the U-Net, PspNet, SegNet, and DeepLabv2 models. Even though this study claims to have worked on small fields, their task focused on cropped versus non-cropped area mapping which led to high accuracy because the labels were highly represented. To enrich the information required for precision agriculture, our research targets crop types instead of cropped areas.

Discouraged by less remote sensing data availability and farming systems at small scales [20], there is not much research done about applying deep learning to satellite data to map crop types in Africa. Despite these challenges, freshly, Rose Rustowicz et al. [21] applied U-Net architecture combined with the CLSTM model to images collected from three satellites: sentinel-1, sentinel-2, and the planet. The model evaluation resulted in an overall accuracy of 60.9% in Ghana.

As leveraging these technologies in Africa continues to be very scarce, our study contributes to the application of state-of-the-art deep learning algorithms to satellite imagery in Ghana. We improve this [21] by using a Swin-UNet architecture on the data collected in Ghana.[3][4]

4 Dataset

The dataset was collected in Ghana to address the scarcity of smallholder farms data. In Africa, smallholder farms typically have smaller fields, resulting in fewer pixels of information[21]. The sparse labels in the dataset introduce gaps in the data. Furthermore, the growing season is affected by rain and cloud cover, which reduces visibility in optical imagery[21].

The dataset was collected in 2016 and 2017 from three different satellites: Sentinel 1, Sentinel 2, and Planet. For Sentinel 2, the spatial resolution is 10m, 30m, and 60m, with 10m covering the visible bands, indicating that 10mx10m is represented in a single pixel. The spectral bands are 10 with 10 days temporal resolution[21]. In the case of Planet, the spatial resolution is an impressive 3 meters, with a revisit time of 1-2 days. This satellite offers four spectral bands: blue, green, red, and infrared. Sentinel 1 provides Synthetic Aperture Radar (SAR) data that is less affected by clouds. It utilizes VV and HH polarization and their ratio to enhance data quality and reduce interference from atmospheric conditions. This feature is particularly valuable for overcoming challenges posed by cloud cover in optical imagery[21].

The dataset comprises 3,466 fields, with 2,259 allocated for training, 298 for validation, and 909 for testing the model. It encompasses 24 crop types, but our focus is on the top four, which collectively represent over 90% of the available crop data. These key crops are Maize (51%), Ground-

[3] Gustave Bwirayesu explored the role of deep learning in remote sensing, emphasizing CNN and the evolution of attention mechanisms in computer vision, introducing the Swin-transformer and its application in Swin-Unet for semantic segmentation

[4] Himbaza B. Jules focused on challenges in crop type mapping with satellite imagery, highlighting efforts to address data scarcity and the global impact of deep learning. He cites examples from Idaho and China, underscoring the scarcity of such research in Africa and positioning their study as a unique contribution using Swin-UNet architecture in Ghana

nut (15%), Rice (14%), and Soya Bean (10%).[4] [5]

5 Baseline

5.1 Baseline Selection

The selection of the baseline was a straightforward process, as we opted for the only existing work conducted on the dataset. The dataset publisher [21] had established a baseline using a 2D U-Net with CLSTM. This decision was made based on the available literature and the recognition of the dataset's unique characteristics for small farms, ensuring a foundation rooted in the established framework.[6]

5.2 Baseline Implementation

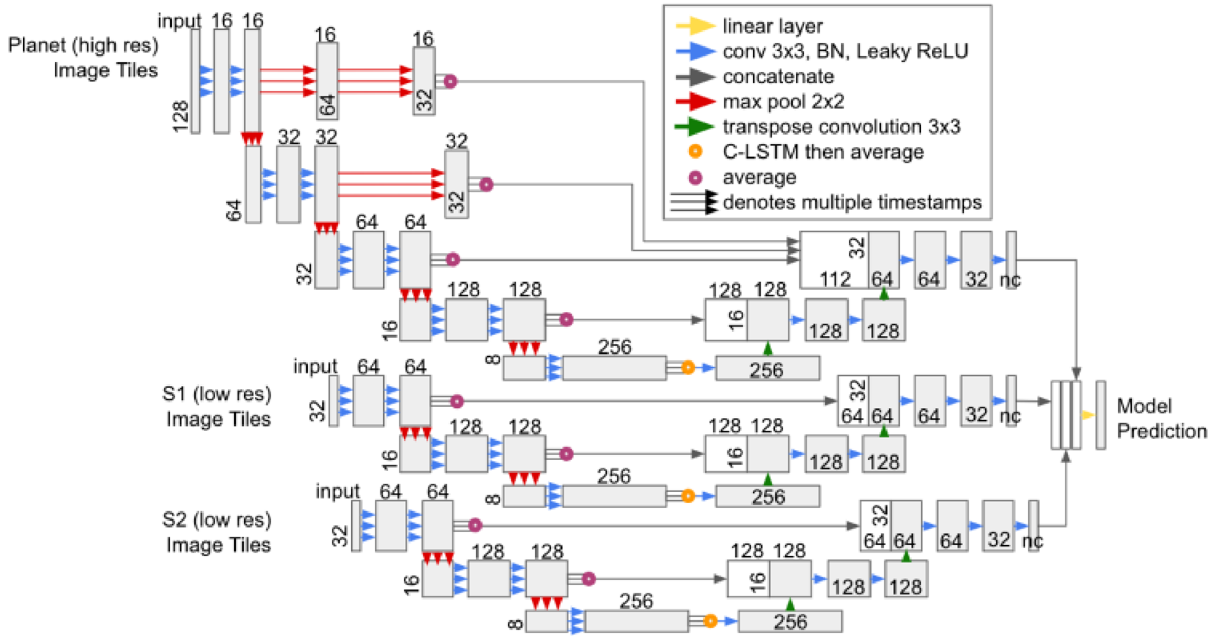


Fig1.The 2D U-Net + CLSTM model architecture used[21]

Rose Rustowicz et al. in [21] studies the problems of small farms in problem in Ghana, South Sudan and Germany. They used U-Net with CLSTM to semantically map the crop types on small farms.In addition to spectral signature, They have used NDVI and GCVI spectral indices to add more context. The Ghana dataset was collected from Sentinel 1, sentinel 2, and planet and each satellite was fed to its UNet model and aggregated their output to produce the final prediction. Their model's evaluation resulted in overall accuracy of 60.9%. Our contribution will focus on trying other state-of-the-art algorithms more specifically Swin-Unet.[7]

[4] Gustave discussed the dataset collection in Ghana, addressing smallholder farm data scarcity and challenges such as sparse labels and reduced visibility due to rain and cloud cover.

[5] Himbaza B. Jules highlights the strategic allocation for training, validation, and testing, providing insights into the dataset's composition and significance for the model

[6] Eric Maniraguha

[7] Deo Uwimpuhwe

6 Methodology

6.1 Model Description

6.1.1 Swin tansformer

In the realm of vision transformers, the Swin Transformer [22] remains a prominent backbone for semantic segmentation tasks. Figure 2 below presents a fairly large Swin-S architecture that was used in our study.

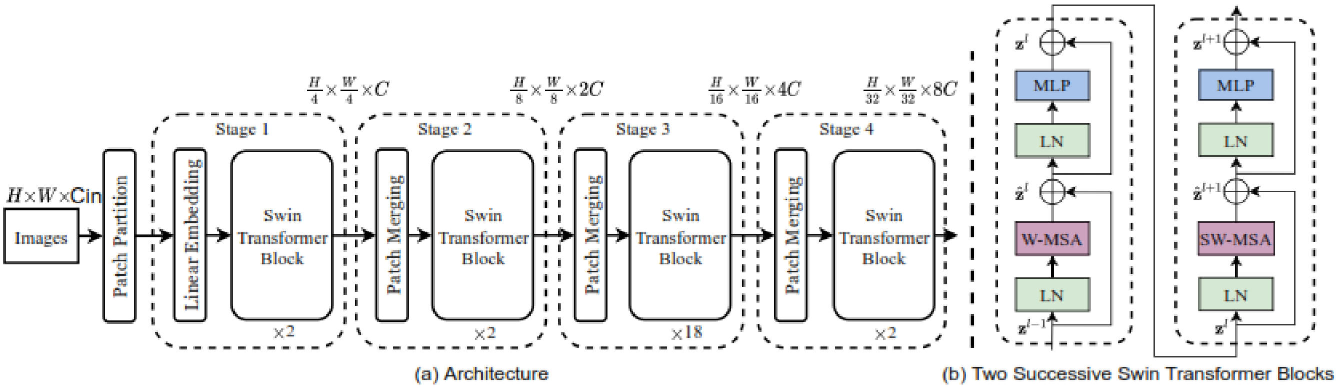


Fig 2. The architecture of the Swin transformer (Swin-S); (a) shows a hierarchical transformation of feature maps in four stages, (b) presents the transformer block with W-MSA and SW-MSA as the key components.

A complete architecture is composed of four stages. At each stage, the feature maps are transformed following the generalized formula
- $N \times \frac{H}{P*2^n} \times \frac{W}{P*2^n} \times C * 2^n$

Where,
n takes the values 0,1,2,3 corresponding to the stages from 1 to 4.
N is the batch size,
H is the height,
W is the width,
P is the patch size, and
C is the number of channels.
The process of patch merging uses a convolutional layer, with kernel size and stride equal to patch size (p), to downsample input images. At each stage, the image is split into non-overlapping windows. Except for stage 3 which has 18 Swin transformer blocks, any other stage has two successive transformer blocks with window multi-head self-attention(W-MSA) and shifted window multi-head self-attention (SW-MSA). The attention is then calculated for the pixel inside the window and window shifting helps in including the neighboring pixels. The formula below which originates from [23] explains the mathematical representation of self-attention.

$Attention(Q, K, V) = SoftMax(\frac{QK^T}{\sqrt{d}} + B) \times V$
where:
Q: Query,
K: key,
V: Values,
d is the dimension of either query or key.

6.1.2 Swin-Unet

The Swin-Unet model [16] is a cutting-edge attention-based architecture designed for image segmentation. It integrates the Swin transformer into a U-net structure, with the encoder part utilizing the Swin transformer. In the decoder part, each stage involves patch expanding, employing the ConvTranspose2d layer and Swin-transformed block. This upsampling process facilitates the application of skip connections between the encoder and decoder, mitigating information loss.

Figure 3 illustrates the complete Swin-Unet architecture implemented for each satellite, while Figure 4 demonstrates the fusion process for the three satellites.
This fusion is achieved by concatenating the output feature maps and passing them through the Conv2D layer to produce feature maps corresponding to the output classes. **appendix 1 presents the model summary of one satellite.** [8][9] [10][11]

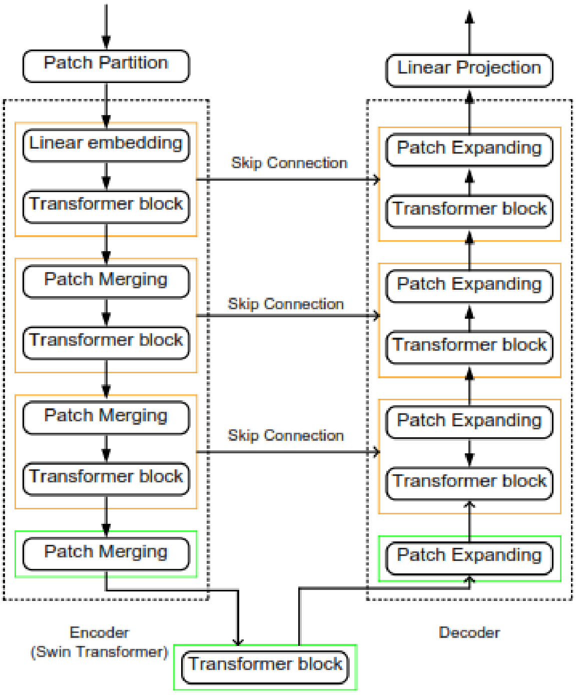
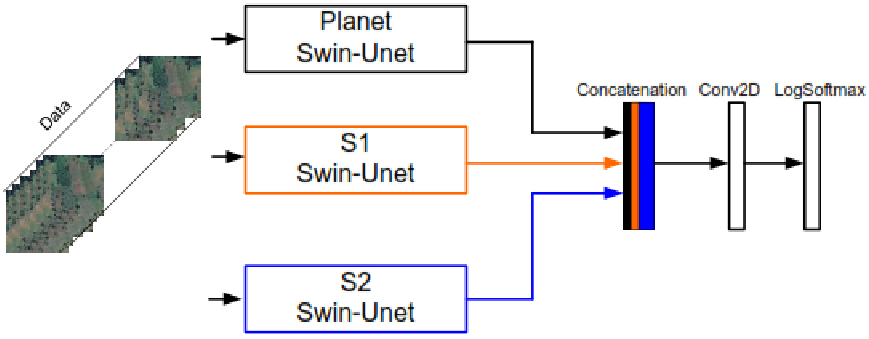


Fig 3. The architecture of the Swin-Unet, encoder uses patch merging for downsampling and the decoder uses patch expanding for upsampling.



[8] Deo Uwimpuhwe
[9] Gustave Bwirayesu
[10] Himbaza B. Jules
[11] Eric Maniraguha

Fig 4. The Swin-Unet architecture that combines three satellites.

6.2 Evaluation metrics

To compare with the baseline, we used overall classification accuracy. This metric gives information on how each pixel is correctly classified; it is computed by considering all the correct classifications over the total data points [24].

6.3 Loss function

In the context of our model, we employ the Negative Log-Likelihood (NLL) loss as the primary objective function. The NLL loss is a widely utilized metric for training classification models, particularly in scenarios where the task involves assigning discrete labels to input data. This loss function provides a clear and quantifiable measure of the disparity between predicted probabilities and actual target distributions.

The NLL loss is computed based on the logarithm of the predicted probabilities assigned to each class [25].

$$NLL\ Loss = -\sum_{i=1}^{N}(y_i \log(f_\Theta(X_i)) + (1 - y_i)\log(1 - f_\Theta(X_i)))$$

Here,
-N is the number of samples,
-y_i is the true label for sample i,
-X_i is the input data.
-f_Θ is the model function

The negative sign is applied to convert the problem of maximizing likelihood into a minimization problem. The loss is high when the predicted probabilities diverge from the true distribution, and it approaches zero when the predicted and true distributions match [25].

The NLL loss is chosen for its suitability in classification problems, aligning with our task of assigning discrete labels to input data. By minimizing this loss during training, our model aims to learn optimal parameters that accurately capture the underlying patterns in the data, leading to improved classification performance. The NLL loss directly addresses the key objective of minimizing the dissimilarity between predicted and true class probabilities.

7 Implemented Experiments

The dataset, structured as a time series with varying time steps across different satellites, prompted us to adopt two distinct approaches. The first involved the random selection of an image within the time dimension, while the second entailed averaging the time series over the growing period. To prepare our data for analysis, we implemented data normalization techniques, addressing intricacies within the dataset for a more effective model evaluation. Additionally, in our continuous quest to enhance model performance, we undertook architectural refinements by varying the number of layers in Swin Transformer configurations. Specifically, we experimented with architectures employing 2, 2, 18, 2 layers (Swin-S) for the four stages[12][13] , and 2, 2, 6, 2 layers (Swin-

[12] Deo Uwimpuhwe

[13] Eric Maniraguha

T), underscoring our dedication to optimizing the model's architecture for superior results. The source code for github repository can be found here or go to the this url: `https://github.com/IntroToDeepLearning-2023` [14][15]

8 Results and Discussions

The results, as per our implementation, include the outputs of two attention-based models (Swin-T and Swin-S). Swin-S performed better than Swin-T. Considering the convergence of the two models, Swin-S reached almost the steady high accuracy at the 40th epochs while Swin-T attained this state at the 45th. Therefore, the results presented herein are the outputs of the Swin-S model. Figure 5 shows the increase of accuracy for 100 epochs and the table 1 compares the overall accuracy of our approach with the baseline.

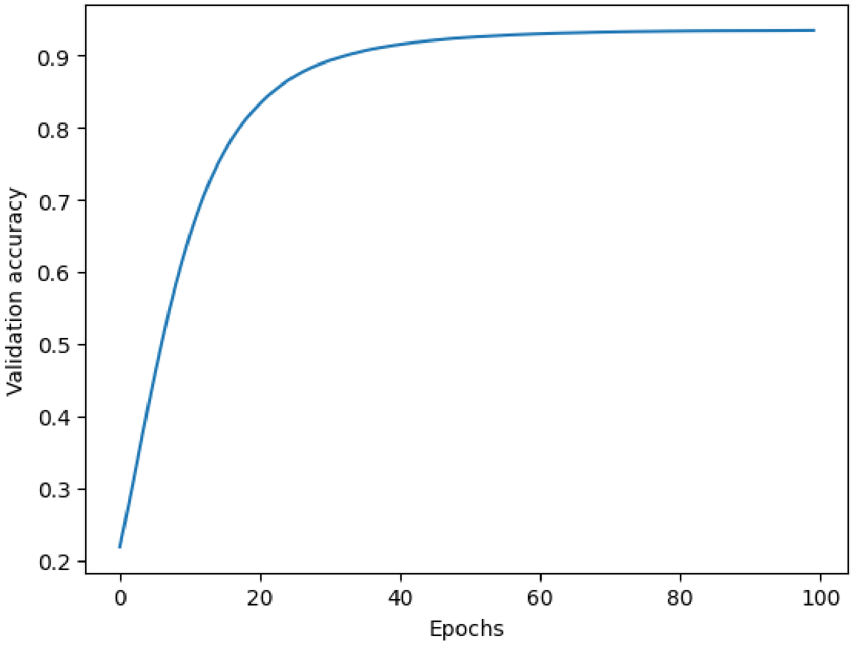


Fig 5. Validation accuracy for 100 epochs.

Performance metric - accuracy	
Performance	Discussion
Our approach: 93.6%	• Achieved higher general accuracy. • Biased on non-cropped areas.
Baseline: 60.9%	• Performed well on classifying crops. • Did not report performance on non-cropped areas.

Table 1. shows the results of the model.

The model achieved a high overall accuracy of 93.6%, a significant improvement over the accu-

[14] Himbaza B. Jules

[15] Gustave Bwirayesu

racy of 60.9% achieved by the related work. This achievement is attributed to the use of Swin-transformer as a backbone which efficiently learns pixel information. Also, averaging time series images contributed to information capture for growing season of crops. The model employed various tuning techniques, such as Layer-wise optimization, to enhance the performance of the model. This helped in optimizing the parameters of the Transformer unit for each satellite dataset. The model showcased promising potential for small fields in Africa. [16][17]

9 Future Works

Although we have made progress in achieving our objective of mapping crop types in small farms, the inherent bias towards non-cropped areas due to the limited size of the cultivated fields underscores the need for additional efforts in developing a more robust model. This undertaking is crucial for effectively addressing the challenges posed by small field sizes and making a meaningful contribution to the progress of precision agriculture in Africa. Our current work, focusing on spectral bands, advocates for the exploration of diverse spectral indices in future research endeavors. To provide richer context, we recommend delving into different spectral indices, and we also suggest investigating various data augmentation techniques. These steps aim to enhance the efficacy and applicability of our model, fostering a more comprehensive and accurate approach to agricultural analysis in the African context.[18]

10 Conclusion

In summary, our work addresses a critical issue – the improvement of agriculture in Africa, a sector that not only employs a significant portion of the population but also contributes substantially to the GDP. Despite the increasing population and persistent hunger, there exists a notable gap in the adoption of advanced technologies, especially in precision agriculture. The prevalent practice of smallholder farming in Africa poses challenges for the application of state-of-the-art deep learning approaches, hindering researchers from contributing to the use of advanced technologies such as satellite data in agriculture. Our approach, utilizing the attention mechanism through Swin-Unet, focused on enhancing performance for smaller agricultural fields and achieved commendable general performance. While we have partially achieved our goal of mapping crop types in small farms, the bias towards non-cropped areas due to small field sizes necessitates further work to develop a robust model capable of addressing these challenges and contributing to the advancement of precision agriculture in Africa.[19]

References

[1] NEPAD. , "Agriculture in Africa: Transformation and oulook," tech. rep.

[2] S. Sakho-Jimbira and I. Hathie, "The future of agriculture in Sub-Saharan Africa," Apr. 2020.

[3] OECD-FAO, "Agriculture in Sub-Saharan Africa: Prospects and challenges for the next decade," agricultural Outlook 2016-2025.

[16] Deo Uwimpuhwe worked on discussion of results

[17] Himbaza B. Jules worked on writing results

[18] Gustave Bwirayesu

[19] Eric Maniraguha

[4] A. O. T. A. UNION, "DECLARATION ON AGRICULTURE AND FOOD SECURITY IN AFRICA," (Maputo, MOZAMBIQUE), July 2003.

[5] A. U. Summit, "Malabo Declaration on Accelerated Agricultural Growth and Transformation for Shared Prosperity and Improved Livelihoods," (Malabo, Equatorial Guinea), June 2014.

[6] F. Z. Bassine, T. E. Epule, A. Kechchour, and A. Chehbouni, "Recent applications of machine learning, remote sensing, and iot approaches in yield prediction: A critical review," June 2023.

[7] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, May 2012.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," Nov. 2020.

[9] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," pp. 166–177, Apr. 2019.

[10] R. Qin and T. Liu, "A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability," 2022.

[11] Z. Li, W. Yang, S. Peng, and F. Liu, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,"

[12] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, July 2022.

[13] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding Deep Learning Techniques for Image Segmentation," July 2019.

[14] Guo Meng-Hao et al., "Attention mechanisms in computer vision: A survey," 2022.

[15] A. Dosovitskiy et al., "AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE," 2021.

[16] Hu Cao et al., "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," May 2021.

[17] "Copernicus Program - an overview — ScienceDirect Topics." https://www.sciencedirect.com/topics/earth-and-planetary-sciences/copernicus-program.

[18] Ehsan Raei et al., "A deep learning image segmentation model for agricultural irrigation system classification,"

[19] Z. Du, J. Yang, C. Ou, and T. Zhang, "Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method," 2019.

[20] Justine M. Nyaga et al., "Precision agriculture research in sub-Saharan Africa countries: A systematic map," Jan. 2021.

[21] Rose Rustowicz et al., “Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods,”

[22] Ze Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” Aug. 2021.

[23] Ashish Vaswani et al., “Attention Is All You Need,” 2017.

[24] M. Grandini, E. Bagli, and G. Visani, “METRICS FOR MULTI-CLASS CLASSIFICATION: AN OVERVIEW,” Aug. 2020.

[25] L. CIAMPICONI et al., “A survey and taxonomy of loss functions in machine learning,” Jan. 2023.

Appendix 1: Swin-Unet architecture for one satellite.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 96, 16, 16]           4,704
 PatchMerging_Conv-2           [-1, 16, 16, 96]               0
            Linear-3          [-1, 16, 16, 288]          27,648
            Linear-4           [-1, 16, 16, 96]           9,312
   WindowAttension-5           [-1, 16, 16, 96]               0
         LayerNorm-6           [-1, 16, 16, 96]             192
           PreNorm-7           [-1, 16, 16, 96]               0
          Residual-8           [-1, 16, 16, 96]               0
            Linear-9          [-1, 16, 16, 384]          37,248
             GELU-10          [-1, 16, 16, 384]               0
           Linear-11           [-1, 16, 16, 96]          36,960
      FeedForward-12           [-1, 16, 16, 96]               0
        LayerNorm-13           [-1, 16, 16, 96]             192
          PreNorm-14           [-1, 16, 16, 96]               0
         Residual-15           [-1, 16, 16, 96]               0
        SwinBlock-16           [-1, 16, 16, 96]               0
      cyclicShift-17           [-1, 16, 16, 96]               0
           Linear-18          [-1, 16, 16, 288]          27,648
           Linear-19           [-1, 16, 16, 96]           9,312
      cyclicShift-20           [-1, 16, 16, 96]               0
  WindowAttension-21           [-1, 16, 16, 96]               0
        LayerNorm-22           [-1, 16, 16, 96]             192
          PreNorm-23           [-1, 16, 16, 96]               0
         Residual-24           [-1, 16, 16, 96]               0
           Linear-25          [-1, 16, 16, 384]          37,248
             GELU-26          [-1, 16, 16, 384]               0
           Linear-27           [-1, 16, 16, 96]          36,960
      FeedForward-28           [-1, 16, 16, 96]               0
        LayerNorm-29           [-1, 16, 16, 96]             192
          PreNorm-30           [-1, 16, 16, 96]               0
         Residual-31           [-1, 16, 16, 96]               0
        SwinBlock-32           [-1, 16, 16, 96]               0
      StageModule-33           [-1, 96, 16, 16]               0
           Conv2d-34            [-1, 192, 8, 8]          73,920
PatchMerging_Conv-35            [-1, 8, 8, 192]               0
           Linear-36            [-1, 8, 8, 576]         110,592
           Linear-37            [-1, 8, 8, 192]          37,056
  WindowAttension-38            [-1, 8, 8, 192]               0
        LayerNorm-39            [-1, 8, 8, 192]             384
          PreNorm-40            [-1, 8, 8, 192]               0
         Residual-41            [-1, 8, 8, 192]               0
           Linear-42            [-1, 8, 8, 768]         148,224
             GELU-43            [-1, 8, 8, 768]               0
           Linear-44            [-1, 8, 8, 192]         147,648
      FeedForward-45            [-1, 8, 8, 192]               0
        LayerNorm-46            [-1, 8, 8, 192]             384
          PreNorm-47            [-1, 8, 8, 192]               0
         Residual-48            [-1, 8, 8, 192]               0
        SwinBlock-49            [-1, 8, 8, 192]               0
      cyclicShift-50            [-1, 8, 8, 192]               0
           Linear-51            [-1, 8, 8, 576]         110,592
           Linear-52            [-1, 8, 8, 192]          37,056
      cyclicShift-53            [-1, 8, 8, 192]               0
  WindowAttension-54            [-1, 8, 8, 192]               0
        LayerNorm-55            [-1, 8, 8, 192]             384
          PreNorm-56            [-1, 8, 8, 192]               0
         Residual-57            [-1, 8, 8, 192]               0
           Linear-58            [-1, 8, 8, 768]         148,224
             GELU-59            [-1, 8, 8, 768]               0
           Linear-60            [-1, 8, 8, 192]         147,648
      FeedForward-61            [-1, 8, 8, 192]               0
        LayerNorm-62            [-1, 8, 8, 192]             384
          PreNorm-63            [-1, 8, 8, 192]               0
         Residual-64            [-1, 8, 8, 192]               0
        SwinBlock-65            [-1, 8, 8, 192]               0
      StageModule-66            [-1, 192, 8, 8]               0
           Conv2d-67            [-1, 384, 4, 4]         295,296
PatchMerging_Conv-68            [-1, 4, 4, 384]               0
           Linear-69           [-1, 4, 4, 1152]         442,368
           Linear-70            [-1, 4, 4, 384]         147,840
  WindowAttension-71            [-1, 4, 4, 384]               0
        LayerNorm-72            [-1, 4, 4, 384]             768
          PreNorm-73            [-1, 4, 4, 384]               0
         Residual-74            [-1, 4, 4, 384]               0
           Linear-75           [-1, 4, 4, 1536]         591,360
             GELU-76           [-1, 4, 4, 1536]               0
           Linear-77            [-1, 4, 4, 384]         590,208
      FeedForward-78            [-1, 4, 4, 384]               0
        LayerNorm-79            [-1, 4, 4, 384]             768
          PreNorm-80            [-1, 4, 4, 384]               0
         Residual-81            [-1, 4, 4, 384]               0
        SwinBlock-82            [-1, 4, 4, 384]               0
      cyclicShift-83            [-1, 4, 4, 384]               0
           Linear-84           [-1, 4, 4, 1152]         442,368
           Linear-85            [-1, 4, 4, 384]         147,840
      cyclicShift-86            [-1, 4, 4, 384]               0
  WindowAttension-87            [-1, 4, 4, 384]               0
        LayerNorm-88            [-1, 4, 4, 384]             768
          PreNorm-89            [-1, 4, 4, 384]               0
         Residual-90            [-1, 4, 4, 384]               0
           Linear-91           [-1, 4, 4, 1536]         591,360
             GELU-92           [-1, 4, 4, 1536]               0
           Linear-93            [-1, 4, 4, 384]         590,208
      FeedForward-94            [-1, 4, 4, 384]               0
        LayerNorm-95            [-1, 4, 4, 384]             768
          PreNorm-96            [-1, 4, 4, 384]               0
         Residual-97            [-1, 4, 4, 384]               0
        SwinBlock-98            [-1, 4, 4, 384]               0
           Linear-99           [-1, 4, 4, 1152]         442,368
          Linear-100            [-1, 4, 4, 384]         147,840
 WindowAttension-101            [-1, 4, 4, 384]               0
       LayerNorm-102            [-1, 4, 4, 384]             768
         PreNorm-103            [-1, 4, 4, 384]               0
        Residual-104            [-1, 4, 4, 384]               0
          Linear-105           [-1, 4, 4, 1536]         591,360
            GELU-106           [-1, 4, 4, 1536]               0
          Linear-107            [-1, 4, 4, 384]         590,208
     FeedForward-108            [-1, 4, 4, 384]               0
       LayerNorm-109            [-1, 4, 4, 384]             768
         PreNorm-110            [-1, 4, 4, 384]               0
        Residual-111            [-1, 4, 4, 384]               0
       SwinBlock-112            [-1, 4, 4, 384]               0
     cyclicShift-113            [-1, 4, 4, 384]               0
          Linear-114           [-1, 4, 4, 1152]         442,368
          Linear-115            [-1, 4, 4, 384]         147,840
     cyclicShift-116            [-1, 4, 4, 384]               0
 WindowAttension-117            [-1, 4, 4, 384]               0
       LayerNorm-118            [-1, 4, 4, 384]             768
         PreNorm-119            [-1, 4, 4, 384]               0
        Residual-120            [-1, 4, 4, 384]               0
          Linear-121           [-1, 4, 4, 1536]         591,360
            GELU-122           [-1, 4, 4, 1536]               0
          Linear-123            [-1, 4, 4, 384]         590,208
     FeedForward-124            [-1, 4, 4, 384]               0
       LayerNorm-125            [-1, 4, 4, 384]             768
         PreNorm-126            [-1, 4, 4, 384]               0
        Residual-127            [-1, 4, 4, 384]               0
       SwinBlock-128            [-1, 4, 4, 384]               0
          Linear-129           [-1, 4, 4, 1152]         442,368
          Linear-130            [-1, 4, 4, 384]         147,840
 WindowAttension-131            [-1, 4, 4, 384]               0
       LayerNorm-132            [-1, 4, 4, 384]             768
         PreNorm-133            [-1, 4, 4, 384]               0
        Residual-134            [-1, 4, 4, 384]               0
          Linear-135           [-1, 4, 4, 1536]         591,360
            GELU-136           [-1, 4, 4, 1536]               0
          Linear-137            [-1, 4, 4, 384]         590,208
     FeedForward-138            [-1, 4, 4, 384]               0
       LayerNorm-139            [-1, 4, 4, 384]             768
         PreNorm-140            [-1, 4, 4, 384]               0
        Residual-141            [-1, 4, 4, 384]               0
       SwinBlock-142            [-1, 4, 4, 384]               0
     cyclicShift-143            [-1, 4, 4, 384]               0
          Linear-144           [-1, 4, 4, 1152]         442,368
          Linear-145            [-1, 4, 4, 384]         147,840
     cyclicShift-146            [-1, 4, 4, 384]               0
 WindowAttension-147            [-1, 4, 4, 384]               0
       LayerNorm-148            [-1, 4, 4, 384]             768
         PreNorm-149            [-1, 4, 4, 384]               0
        Residual-150            [-1, 4, 4, 384]               0
          Linear-151           [-1, 4, 4, 1536]         591,360
            GELU-152           [-1, 4, 4, 1536]               0
          Linear-153            [-1, 4, 4, 384]         590,208
     FeedForward-154            [-1, 4, 4, 384]               0
       LayerNorm-155            [-1, 4, 4, 384]             768
         PreNorm-156            [-1, 4, 4, 384]               0
        Residual-157            [-1, 4, 4, 384]               0
       SwinBlock-158            [-1, 4, 4, 384]               0
          Linear-159           [-1, 4, 4, 1152]         442,368
          Linear-160            [-1, 4, 4, 384]         147,840
 WindowAttension-161            [-1, 4, 4, 384]               0
       LayerNorm-162            [-1, 4, 4, 384]             768
         PreNorm-163            [-1, 4, 4, 384]               0
        Residual-164            [-1, 4, 4, 384]               0
          Linear-165           [-1, 4, 4, 1536]         591,360
            GELU-166           [-1, 4, 4, 1536]               0
          Linear-167            [-1, 4, 4, 384]         590,208
     FeedForward-168            [-1, 4, 4, 384]               0
       LayerNorm-169            [-1, 4, 4, 384]             768
```

```
         PreNorm-170            [-1, 4, 4, 384]               0
        Residual-171            [-1, 4, 4, 384]               0
       SwinBlock-172            [-1, 4, 4, 384]               0
     cyclicShift-173            [-1, 4, 4, 384]               0
          Linear-174           [-1, 4, 4, 1152]         442,368
          Linear-175            [-1, 4, 4, 384]         147,840
     cyclicShift-176            [-1, 4, 4, 384]               0
 WindowAttention-177            [-1, 4, 4, 384]               0
       LayerNorm-178            [-1, 4, 4, 384]             768
         PreNorm-179            [-1, 4, 4, 384]               0
        Residual-180            [-1, 4, 4, 384]               0
          Linear-181           [-1, 4, 4, 1536]         591,360
            GELU-182           [-1, 4, 4, 1536]               0
          Linear-183            [-1, 4, 4, 384]         590,208
     FeedForward-184            [-1, 4, 4, 384]               0
       LayerNorm-185            [-1, 4, 4, 384]             768
         PreNorm-186            [-1, 4, 4, 384]               0
        Residual-187            [-1, 4, 4, 384]               0
       SwinBlock-188            [-1, 4, 4, 384]               0
          Linear-189           [-1, 4, 4, 1152]         442,368
          Linear-190            [-1, 4, 4, 384]         147,840
 WindowAttention-191            [-1, 4, 4, 384]               0
       LayerNorm-192            [-1, 4, 4, 384]             768
         PreNorm-193            [-1, 4, 4, 384]               0
        Residual-194            [-1, 4, 4, 384]               0
          Linear-195           [-1, 4, 4, 1536]         591,360
            GELU-196           [-1, 4, 4, 1536]               0
          Linear-197            [-1, 4, 4, 384]         590,208
     FeedForward-198            [-1, 4, 4, 384]               0
       LayerNorm-199            [-1, 4, 4, 384]             768
         PreNorm-200            [-1, 4, 4, 384]               0
        Residual-201            [-1, 4, 4, 384]               0
       SwinBlock-202            [-1, 4, 4, 384]               0
     cyclicShift-203            [-1, 4, 4, 384]               0
          Linear-204           [-1, 4, 4, 1152]         442,368
          Linear-205            [-1, 4, 4, 384]         147,840
     cyclicShift-206            [-1, 4, 4, 384]               0
 WindowAttention-207            [-1, 4, 4, 384]               0
       LayerNorm-208            [-1, 4, 4, 384]             768
         PreNorm-209            [-1, 4, 4, 384]               0
        Residual-210            [-1, 4, 4, 384]               0
          Linear-211           [-1, 4, 4, 1536]         591,360
            GELU-212           [-1, 4, 4, 1536]               0
          Linear-213            [-1, 4, 4, 384]         590,208
     FeedForward-214            [-1, 4, 4, 384]               0
       LayerNorm-215            [-1, 4, 4, 384]             768
         PreNorm-216            [-1, 4, 4, 384]               0
        Residual-217            [-1, 4, 4, 384]               0
       SwinBlock-218            [-1, 4, 4, 384]               0
          Linear-219           [-1, 4, 4, 1152]         442,368
          Linear-220            [-1, 4, 4, 384]         147,840
 WindowAttention-221            [-1, 4, 4, 384]               0
       LayerNorm-222            [-1, 4, 4, 384]             768
         PreNorm-223            [-1, 4, 4, 384]               0
        Residual-224            [-1, 4, 4, 384]               0
          Linear-225           [-1, 4, 4, 1536]         591,360
            GELU-226           [-1, 4, 4, 1536]               0
          Linear-227            [-1, 4, 4, 384]         590,208
     FeedForward-228            [-1, 4, 4, 384]               0
       LayerNorm-229            [-1, 4, 4, 384]             768
         PreNorm-230            [-1, 4, 4, 384]               0
        Residual-231            [-1, 4, 4, 384]               0
       SwinBlock-232            [-1, 4, 4, 384]               0
     cyclicShift-233            [-1, 4, 4, 384]               0
          Linear-234           [-1, 4, 4, 1152]         442,368
          Linear-235            [-1, 4, 4, 384]         147,840
     cyclicShift-236            [-1, 4, 4, 384]               0
 WindowAttension-237            [-1, 4, 4, 384]               0
       LayerNorm-238            [-1, 4, 4, 384]             768
         PreNorm-239            [-1, 4, 4, 384]               0
        Residual-240            [-1, 4, 4, 384]               0
          Linear-241           [-1, 4, 4, 1536]         591,360
            GELU-242           [-1, 4, 4, 1536]               0
          Linear-243            [-1, 4, 4, 384]         590,208
     FeedForward-244            [-1, 4, 4, 384]               0
       LayerNorm-245            [-1, 4, 4, 384]             768
         PreNorm-246            [-1, 4, 4, 384]               0
        Residual-247            [-1, 4, 4, 384]               0
       SwinBlock-248            [-1, 4, 4, 384]               0
          Linear-249           [-1, 4, 4, 1152]         442,368
          Linear-250            [-1, 4, 4, 384]         147,840
 WindowAttension-251            [-1, 4, 4, 384]               0
       LayerNorm-252            [-1, 4, 4, 384]             768
         PreNorm-253            [-1, 4, 4, 384]               0
        Residual-254            [-1, 4, 4, 384]               0
          Linear-255           [-1, 4, 4, 1536]         591,360
            GELU-256           [-1, 4, 4, 1536]               0
          Linear-257            [-1, 4, 4, 384]         590,208
     FeedForward-258            [-1, 4, 4, 384]               0
       LayerNorm-259            [-1, 4, 4, 384]             768
         PreNorm-260            [-1, 4, 4, 384]               0
        Residual-261            [-1, 4, 4, 384]               0
       SwinBlock-262            [-1, 4, 4, 384]               0
     cyclicShift-263            [-1, 4, 4, 384]               0
          Linear-264           [-1, 4, 4, 1152]         442,368
          Linear-265            [-1, 4, 4, 384]         147,840
     cyclicShift-266            [-1, 4, 4, 384]               0
 WindowAttension-267            [-1, 4, 4, 384]               0
       LayerNorm-268            [-1, 4, 4, 384]             768
         PreNorm-269            [-1, 4, 4, 384]               0
        Residual-270            [-1, 4, 4, 384]               0
          Linear-271           [-1, 4, 4, 1536]         591,360
            GELU-272           [-1, 4, 4, 1536]               0
          Linear-273            [-1, 4, 4, 384]         590,208
     FeedForward-274            [-1, 4, 4, 384]               0
       LayerNorm-275            [-1, 4, 4, 384]             768
         PreNorm-276            [-1, 4, 4, 384]               0
        Residual-277            [-1, 4, 4, 384]               0
       SwinBlock-278            [-1, 4, 4, 384]               0
          Linear-279           [-1, 4, 4, 1152]         442,368
          Linear-280            [-1, 4, 4, 384]         147,840
 WindowAttension-281            [-1, 4, 4, 384]               0
       LayerNorm-282            [-1, 4, 4, 384]             768
         PreNorm-283            [-1, 4, 4, 384]               0
        Residual-284            [-1, 4, 4, 384]               0
          Linear-285           [-1, 4, 4, 1536]         591,360
```

```
            GELU-286           [-1, 4, 4, 1536]               0
          Linear-287            [-1, 4, 4, 384]         590,208
     FeedForward-288            [-1, 4, 4, 384]               0
       LayerNorm-289            [-1, 4, 4, 384]             768
         PreNorm-290            [-1, 4, 4, 384]               0
        Residual-291            [-1, 4, 4, 384]               0
       SwinBlock-292            [-1, 4, 4, 384]               0
     cyclicShift-293            [-1, 4, 4, 384]               0
          Linear-294           [-1, 4, 4, 1152]         442,368
          Linear-295            [-1, 4, 4, 384]         147,840
     cyclicShift-296            [-1, 4, 4, 384]               0
 WindowAttention-297            [-1, 4, 4, 384]               0
       LayerNorm-298            [-1, 4, 4, 384]             768
         PreNorm-299            [-1, 4, 4, 384]               0
        Residual-300            [-1, 4, 4, 384]               0
          Linear-301           [-1, 4, 4, 1536]         591,360
            GELU-302           [-1, 4, 4, 1536]               0
          Linear-303            [-1, 4, 4, 384]         590,208
     FeedForward-304            [-1, 4, 4, 384]               0
       LayerNorm-305            [-1, 4, 4, 384]             768
         PreNorm-306            [-1, 4, 4, 384]               0
        Residual-307            [-1, 4, 4, 384]               0
       SwinBlock-308            [-1, 4, 4, 384]               0
          Linear-309           [-1, 4, 4, 1152]         442,368
          Linear-310            [-1, 4, 4, 384]         147,840
 WindowAttention-311            [-1, 4, 4, 384]               0
       LayerNorm-312            [-1, 4, 4, 384]             768
         PreNorm-313            [-1, 4, 4, 384]               0
        Residual-314            [-1, 4, 4, 384]               0
          Linear-315           [-1, 4, 4, 1536]         591,360
            GELU-316           [-1, 4, 4, 1536]               0
          Linear-317            [-1, 4, 4, 384]         590,208
     FeedForward-318            [-1, 4, 4, 384]               0
       LayerNorm-319            [-1, 4, 4, 384]             768
         PreNorm-320            [-1, 4, 4, 384]               0
        Residual-321            [-1, 4, 4, 384]               0
       SwinBlock-322            [-1, 4, 4, 384]               0
     cyclicShift-323            [-1, 4, 4, 384]               0
          Linear-324           [-1, 4, 4, 1152]         442,368
          Linear-325            [-1, 4, 4, 384]         147,840
     cyclicShift-326            [-1, 4, 4, 384]               0
 WindowAttention-327            [-1, 4, 4, 384]               0
       LayerNorm-328            [-1, 4, 4, 384]             768
         PreNorm-329            [-1, 4, 4, 384]               0
        Residual-330            [-1, 4, 4, 384]               0
          Linear-331           [-1, 4, 4, 1536]         591,360
            GELU-332           [-1, 4, 4, 1536]               0
          Linear-333            [-1, 4, 4, 384]         590,208
     FeedForward-334            [-1, 4, 4, 384]               0
       LayerNorm-335            [-1, 4, 4, 384]             768
         PreNorm-336            [-1, 4, 4, 384]               0
        Residual-337            [-1, 4, 4, 384]               0
       SwinBlock-338            [-1, 4, 4, 384]               0
     StageModule-339            [-1, 384, 4, 4]               0
          Conv2d-340            [-1, 768, 2, 2]       1,180,416
PatchMerging_Conv-341            [-1, 2, 2, 768]               0
          Linear-342           [-1, 2, 2, 2304]       1,769,472
          Linear-343            [-1, 2, 2, 768]         590,592
 WindowAttention-344            [-1, 2, 2, 768]               0
       LayerNorm-345            [-1, 2, 2, 768]           1,536
         PreNorm-346            [-1, 2, 2, 768]               0
        Residual-347            [-1, 2, 2, 768]               0
          Linear-348           [-1, 2, 2, 3072]       2,362,368
            GELU-349           [-1, 2, 2, 3072]               0
          Linear-350            [-1, 2, 2, 768]       2,360,064
     FeedForward-351            [-1, 2, 2, 768]               0
       LayerNorm-352            [-1, 2, 2, 768]           1,536
         PreNorm-353            [-1, 2, 2, 768]               0
        Residual-354            [-1, 2, 2, 768]               0
       SwinBlock-355            [-1, 2, 2, 768]               0
     cyclicShift-356            [-1, 2, 2, 768]               0
          Linear-357           [-1, 2, 2, 2304]       1,769,472
          Linear-358            [-1, 2, 2, 768]         590,592
     cyclicShift-359            [-1, 2, 2, 768]               0
 WindowAttention-360            [-1, 2, 2, 768]               0
       LayerNorm-361            [-1, 2, 2, 768]           1,536
         PreNorm-362            [-1, 2, 2, 768]               0
        Residual-363            [-1, 2, 2, 768]               0
          Linear-364           [-1, 2, 2, 3072]       2,362,368
            GELU-365           [-1, 2, 2, 3072]               0
          Linear-366            [-1, 2, 2, 768]       2,360,064
     FeedForward-367            [-1, 2, 2, 768]               0
       LayerNorm-368            [-1, 2, 2, 768]           1,536
         PreNorm-369            [-1, 2, 2, 768]               0
        Residual-370            [-1, 2, 2, 768]               0
       SwinBlock-371            [-1, 2, 2, 768]               0
     StageModule-372            [-1, 768, 2, 2]               0
 ConvTranspose2d-373            [-1, 384, 4, 4]       2,654,592
  PatchExpanding-374            [-1, 4, 4, 384]               0
          Linear-375           [-1, 4, 4, 2304]         884,736
          Linear-376            [-1, 4, 4, 384]         295,296
 WindowAttention-377            [-1, 4, 4, 384]               0
       LayerNorm-378            [-1, 4, 4, 384]             768
         PreNorm-379            [-1, 4, 4, 384]               0
        Residual-380            [-1, 4, 4, 384]               0
          Linear-381           [-1, 4, 4, 1536]         591,360
            GELU-382           [-1, 4, 4, 1536]               0
          Linear-383            [-1, 4, 4, 384]         590,208
     FeedForward-384            [-1, 4, 4, 384]               0
       LayerNorm-385            [-1, 4, 4, 384]             768
         PreNorm-386            [-1, 4, 4, 384]               0
        Residual-387            [-1, 4, 4, 384]               0
       SwinBlock-388            [-1, 4, 4, 384]               0
     cyclicShift-389            [-1, 4, 4, 384]               0
          Linear-390           [-1, 4, 4, 2304]         884,736
          Linear-391            [-1, 4, 4, 384]         295,296
     cyclicShift-392            [-1, 4, 4, 384]               0
 WindowAttention-393            [-1, 4, 4, 384]               0
       LayerNorm-394            [-1, 4, 4, 384]             768
         PreNorm-395            [-1, 4, 4, 384]               0
        Residual-396            [-1, 4, 4, 384]               0
          Linear-397           [-1, 4, 4, 1536]         591,360
            GELU-398           [-1, 4, 4, 1536]               0
          Linear-399            [-1, 4, 4, 384]         590,208
     FeedForward-400            [-1, 4, 4, 384]               0
       LayerNorm-401            [-1, 4, 4, 384]             768
```

```
         PreNorm-402            [-1, 4, 4, 384]               0
        Residual-403            [-1, 4, 4, 384]               0
       SwinBlock-404            [-1, 4, 4, 384]               0
     StageModule-405            [-1, 384, 4, 4]               0
 ConvTranspose2d-406            [-1, 192, 8, 8]       1,327,296
  PatchExpanding-407            [-1, 8, 8, 192]               0
          Linear-408           [-1, 8, 8, 1152]         221,184
          Linear-409            [-1, 8, 8, 192]          73,920
 WindowAttension-410            [-1, 8, 8, 192]               0
       LayerNorm-411            [-1, 8, 8, 192]             384
         PreNorm-412            [-1, 8, 8, 192]               0
        Residual-413            [-1, 8, 8, 192]               0
          Linear-414            [-1, 8, 8, 768]         148,224
            GELU-415            [-1, 8, 8, 768]               0
          Linear-416            [-1, 8, 8, 192]         147,648
     FeedForward-417            [-1, 8, 8, 192]               0
       LayerNorm-418            [-1, 8, 8, 192]             384
         PreNorm-419            [-1, 8, 8, 192]               0
        Residual-420            [-1, 8, 8, 192]               0
       SwinBlock-421            [-1, 8, 8, 192]               0
     cyclicShift-422            [-1, 8, 8, 192]               0
          Linear-423           [-1, 8, 8, 1152]         221,184
          Linear-424            [-1, 8, 8, 192]          73,920
     cyclicShift-425            [-1, 8, 8, 192]               0
 WindowAttension-426            [-1, 8, 8, 192]               0
       LayerNorm-427            [-1, 8, 8, 192]             384
         PreNorm-428            [-1, 8, 8, 192]               0
        Residual-429            [-1, 8, 8, 192]               0
          Linear-430            [-1, 8, 8, 768]         148,224
            GELU-431            [-1, 8, 8, 768]               0
          Linear-432            [-1, 8, 8, 192]         147,648
     FeedForward-433            [-1, 8, 8, 192]               0
       LayerNorm-434            [-1, 8, 8, 192]             384
         PreNorm-435            [-1, 8, 8, 192]               0
        Residual-436            [-1, 8, 8, 192]               0
       SwinBlock-437            [-1, 8, 8, 192]               0
     StageModule-438            [-1, 192, 8, 8]               0
 ConvTranspose2d-439           [-1, 96, 16, 16]         331,872
  PatchExpanding-440           [-1, 16, 16, 96]               0
          Linear-441          [-1, 16, 16, 576]          55,296
          Linear-442           [-1, 16, 16, 96]          18,528
 WindowAttension-443           [-1, 16, 16, 96]               0
       LayerNorm-444           [-1, 16, 16, 96]             192
         PreNorm-445           [-1, 16, 16, 96]               0
        Residual-446           [-1, 16, 16, 96]               0
          Linear-447          [-1, 16, 16, 384]          37,248
            GELU-448          [-1, 16, 16, 384]               0
          Linear-449           [-1, 16, 16, 96]          36,960
     FeedForward-450           [-1, 16, 16, 96]               0
       LayerNorm-451           [-1, 16, 16, 96]             192
         PreNorm-452           [-1, 16, 16, 96]               0
        Residual-453           [-1, 16, 16, 96]               0
       SwinBlock-454           [-1, 16, 16, 96]               0
     cyclicShift-455           [-1, 16, 16, 96]               0
          Linear-456          [-1, 16, 16, 576]          55,296
          Linear-457           [-1, 16, 16, 96]          18,528
     cyclicShift-458           [-1, 16, 16, 96]               0
 WindowAttension-459           [-1, 16, 16, 96]               0
```

```
       LayerNorm-460           [-1, 16, 16, 96]             192
         PreNorm-461           [-1, 16, 16, 96]               0
        Residual-462           [-1, 16, 16, 96]               0
          Linear-463          [-1, 16, 16, 384]          37,248
            GELU-464          [-1, 16, 16, 384]               0
          Linear-465           [-1, 16, 16, 96]          36,960
     FeedForward-466           [-1, 16, 16, 96]               0
       LayerNorm-467           [-1, 16, 16, 96]             192
         PreNorm-468           [-1, 16, 16, 96]               0
        Residual-469           [-1, 16, 16, 96]               0
       SwinBlock-470           [-1, 16, 16, 96]               0
          Linear-471          [-1, 16, 16, 576]          55,296
          Linear-472           [-1, 16, 16, 96]          18,528
 WindowAttension-473           [-1, 16, 16, 96]               0
       LayerNorm-474           [-1, 16, 16, 96]             192
         PreNorm-475           [-1, 16, 16, 96]               0
        Residual-476           [-1, 16, 16, 96]               0
          Linear-477          [-1, 16, 16, 384]          37,248
            GELU-478          [-1, 16, 16, 384]               0
          Linear-479           [-1, 16, 16, 96]          36,960
     FeedForward-480           [-1, 16, 16, 96]               0
       LayerNorm-481           [-1, 16, 16, 96]             192
         PreNorm-482           [-1, 16, 16, 96]               0
        Residual-483           [-1, 16, 16, 96]               0
       SwinBlock-484           [-1, 16, 16, 96]               0
     cyclicShift-485           [-1, 16, 16, 96]               0
          Linear-486          [-1, 16, 16, 576]          55,296
          Linear-487           [-1, 16, 16, 96]          18,528
     cyclicShift-488           [-1, 16, 16, 96]               0
 WindowAttension-489           [-1, 16, 16, 96]               0
       LayerNorm-490           [-1, 16, 16, 96]             192
         PreNorm-491           [-1, 16, 16, 96]               0
        Residual-492           [-1, 16, 16, 96]               0
          Linear-493          [-1, 16, 16, 384]          37,248
            GELU-494          [-1, 16, 16, 384]               0
          Linear-495           [-1, 16, 16, 96]          36,960
     FeedForward-496           [-1, 16, 16, 96]               0
       LayerNorm-497           [-1, 16, 16, 96]             192
         PreNorm-498           [-1, 16, 16, 96]               0
        Residual-499           [-1, 16, 16, 96]               0
       SwinBlock-500           [-1, 16, 16, 96]               0
          Linear-501          [-1, 16, 16, 576]          55,296
          Linear-502           [-1, 16, 16, 96]          18,528
 WindowAttension-503           [-1, 16, 16, 96]               0
       LayerNorm-504           [-1, 16, 16, 96]             192
         PreNorm-505           [-1, 16, 16, 96]               0
        Residual-506           [-1, 16, 16, 96]               0
          Linear-507          [-1, 16, 16, 384]          37,248
            GELU-508          [-1, 16, 16, 384]               0
          Linear-509           [-1, 16, 16, 96]          36,960
     FeedForward-510           [-1, 16, 16, 96]               0
       LayerNorm-511           [-1, 16, 16, 96]             192
         PreNorm-512           [-1, 16, 16, 96]               0
        Residual-513           [-1, 16, 16, 96]               0
       SwinBlock-514           [-1, 16, 16, 96]               0
     cyclicShift-515           [-1, 16, 16, 96]               0
          Linear-516          [-1, 16, 16, 576]          55,296
          Linear-517           [-1, 16, 16, 96]          18,528
     cyclicShift-518           [-1, 16, 16, 96]               0
 WindowAttension-519           [-1, 16, 16, 96]               0
       LayerNorm-520           [-1, 16, 16, 96]             192
         PreNorm-521           [-1, 16, 16, 96]               0
        Residual-522           [-1, 16, 16, 96]               0
          Linear-523          [-1, 16, 16, 384]          37,248
            GELU-524          [-1, 16, 16, 384]               0
          Linear-525           [-1, 16, 16, 96]          36,960
     FeedForward-526           [-1, 16, 16, 96]               0
       LayerNorm-527           [-1, 16, 16, 96]             192
         PreNorm-528           [-1, 16, 16, 96]               0
        Residual-529           [-1, 16, 16, 96]               0
       SwinBlock-530           [-1, 16, 16, 96]               0
          Linear-531          [-1, 16, 16, 576]          55,296
          Linear-532           [-1, 16, 16, 96]          18,528
 WindowAttension-533           [-1, 16, 16, 96]               0
       LayerNorm-534           [-1, 16, 16, 96]             192
         PreNorm-535           [-1, 16, 16, 96]               0
        Residual-536           [-1, 16, 16, 96]               0
          Linear-537          [-1, 16, 16, 384]          37,248
            GELU-538          [-1, 16, 16, 384]               0
          Linear-539           [-1, 16, 16, 96]          36,960
     FeedForward-540           [-1, 16, 16, 96]               0
       LayerNorm-541           [-1, 16, 16, 96]             192
         PreNorm-542           [-1, 16, 16, 96]               0
        Residual-543           [-1, 16, 16, 96]               0
       SwinBlock-544           [-1, 16, 16, 96]               0
     cyclicShift-545           [-1, 16, 16, 96]               0
          Linear-546          [-1, 16, 16, 576]          55,296
          Linear-547           [-1, 16, 16, 96]          18,528
     cyclicShift-548           [-1, 16, 16, 96]               0
 WindowAttension-549           [-1, 16, 16, 96]               0
       LayerNorm-550           [-1, 16, 16, 96]             192
         PreNorm-551           [-1, 16, 16, 96]               0
        Residual-552           [-1, 16, 16, 96]               0
          Linear-553          [-1, 16, 16, 384]          37,248
            GELU-554          [-1, 16, 16, 384]               0
          Linear-555           [-1, 16, 16, 96]          36,960
     FeedForward-556           [-1, 16, 16, 96]               0
       LayerNorm-557           [-1, 16, 16, 96]             192
         PreNorm-558           [-1, 16, 16, 96]               0
        Residual-559           [-1, 16, 16, 96]               0
       SwinBlock-560           [-1, 16, 16, 96]               0
          Linear-561          [-1, 16, 16, 576]          55,296
          Linear-562           [-1, 16, 16, 96]          18,528
 WindowAttension-563           [-1, 16, 16, 96]               0
       LayerNorm-564           [-1, 16, 16, 96]             192
         PreNorm-565           [-1, 16, 16, 96]               0
        Residual-566           [-1, 16, 16, 96]               0
          Linear-567          [-1, 16, 16, 384]          37,248
            GELU-568          [-1, 16, 16, 384]               0
          Linear-569           [-1, 16, 16, 96]          36,960
     FeedForward-570           [-1, 16, 16, 96]               0
       LayerNorm-571           [-1, 16, 16, 96]             192
         PreNorm-572           [-1, 16, 16, 96]               0
        Residual-573           [-1, 16, 16, 96]               0
       SwinBlock-574           [-1, 16, 16, 96]               0
     cyclicShift-575           [-1, 16, 16, 96]               0
```

```
          Linear-576          [-1, 16, 16, 576]          55,296
          Linear-577           [-1, 16, 16, 96]          18,528
     cyclicShift-578           [-1, 16, 16, 96]               0
 WindowAttension-579           [-1, 16, 16, 96]               0
       LayerNorm-580           [-1, 16, 16, 96]             192
         PreNorm-581           [-1, 16, 16, 96]               0
        Residual-582           [-1, 16, 16, 96]               0
          Linear-583          [-1, 16, 16, 384]          37,248
            GELU-584          [-1, 16, 16, 384]               0
          Linear-585           [-1, 16, 16, 96]          36,960
     FeedForward-586           [-1, 16, 16, 96]               0
       LayerNorm-587           [-1, 16, 16, 96]             192
         PreNorm-588           [-1, 16, 16, 96]               0
        Residual-589           [-1, 16, 16, 96]               0
       SwinBlock-590           [-1, 16, 16, 96]               0
          Linear-591          [-1, 16, 16, 576]          55,296
          Linear-592           [-1, 16, 16, 96]          18,528
 WindowAttension-593           [-1, 16, 16, 96]               0
       LayerNorm-594           [-1, 16, 16, 96]             192
         PreNorm-595           [-1, 16, 16, 96]               0
        Residual-596           [-1, 16, 16, 96]               0
          Linear-597          [-1, 16, 16, 384]          37,248
            GELU-598          [-1, 16, 16, 384]               0
          Linear-599           [-1, 16, 16, 96]          36,960
     FeedForward-600           [-1, 16, 16, 96]               0
       LayerNorm-601           [-1, 16, 16, 96]             192
         PreNorm-602           [-1, 16, 16, 96]               0
        Residual-603           [-1, 16, 16, 96]               0
       SwinBlock-604           [-1, 16, 16, 96]               0
     cyclicShift-605           [-1, 16, 16, 96]               0
          Linear-606          [-1, 16, 16, 576]          55,296
          Linear-607           [-1, 16, 16, 96]          18,528
     cyclicShift-608           [-1, 16, 16, 96]               0
 WindowAttension-609           [-1, 16, 16, 96]               0
       LayerNorm-610           [-1, 16, 16, 96]             192
         PreNorm-611           [-1, 16, 16, 96]               0
        Residual-612           [-1, 16, 16, 96]               0
          Linear-613          [-1, 16, 16, 384]          37,248
            GELU-614          [-1, 16, 16, 384]               0
          Linear-615           [-1, 16, 16, 96]          36,960
     FeedForward-616           [-1, 16, 16, 96]               0
       LayerNorm-617           [-1, 16, 16, 96]             192
         PreNorm-618           [-1, 16, 16, 96]               0
        Residual-619           [-1, 16, 16, 96]               0
       SwinBlock-620           [-1, 16, 16, 96]               0
          Linear-621          [-1, 16, 16, 576]          55,296
          Linear-622           [-1, 16, 16, 96]          18,528
 WindowAttension-623           [-1, 16, 16, 96]               0
       LayerNorm-624           [-1, 16, 16, 96]             192
         PreNorm-625           [-1, 16, 16, 96]               0
        Residual-626           [-1, 16, 16, 96]               0
          Linear-627          [-1, 16, 16, 384]          37,248
            GELU-628          [-1, 16, 16, 384]               0
          Linear-629           [-1, 16, 16, 96]          36,960
     FeedForward-630           [-1, 16, 16, 96]               0
       LayerNorm-631           [-1, 16, 16, 96]             192
         PreNorm-632           [-1, 16, 16, 96]               0
        Residual-633           [-1, 16, 16, 96]               0
       SwinBlock-634           [-1, 16, 16, 96]               0
     cyclicShift-635           [-1, 16, 16, 96]               0
          Linear-636          [-1, 16, 16, 576]          55,296
          Linear-637           [-1, 16, 16, 96]          18,528
     cyclicShift-638           [-1, 16, 16, 96]               0
 WindowAttention-639           [-1, 16, 16, 96]               0
       LayerNorm-640           [-1, 16, 16, 96]             192
         PreNorm-641           [-1, 16, 16, 96]               0
        Residual-642           [-1, 16, 16, 96]               0
          Linear-643          [-1, 16, 16, 384]          37,248
            GELU-644          [-1, 16, 16, 384]               0
          Linear-645           [-1, 16, 16, 96]          36,960
     FeedForward-646           [-1, 16, 16, 96]               0
       LayerNorm-647           [-1, 16, 16, 96]             192
         PreNorm-648           [-1, 16, 16, 96]               0
        Residual-649           [-1, 16, 16, 96]               0
       SwinBlock-650           [-1, 16, 16, 96]               0
          Linear-651          [-1, 16, 16, 576]          55,296
          Linear-652           [-1, 16, 16, 96]          18,528
 WindowAttention-653           [-1, 16, 16, 96]               0
       LayerNorm-654           [-1, 16, 16, 96]             192
         PreNorm-655           [-1, 16, 16, 96]               0
        Residual-656           [-1, 16, 16, 96]               0
          Linear-657          [-1, 16, 16, 384]          37,248
            GELU-658          [-1, 16, 16, 384]               0
          Linear-659           [-1, 16, 16, 96]          36,960
     FeedForward-660           [-1, 16, 16, 96]               0
       LayerNorm-661           [-1, 16, 16, 96]             192
         PreNorm-662           [-1, 16, 16, 96]               0
        Residual-663           [-1, 16, 16, 96]               0
       SwinBlock-664           [-1, 16, 16, 96]               0
     cyclicShift-665           [-1, 16, 16, 96]               0
          Linear-666          [-1, 16, 16, 576]          55,296
          Linear-667           [-1, 16, 16, 96]          18,528
     cyclicShift-668           [-1, 16, 16, 96]               0
 WindowAttention-669           [-1, 16, 16, 96]               0
       LayerNorm-670           [-1, 16, 16, 96]             192
         PreNorm-671           [-1, 16, 16, 96]               0
        Residual-672           [-1, 16, 16, 96]               0
          Linear-673          [-1, 16, 16, 384]          37,248
            GELU-674          [-1, 16, 16, 384]               0
          Linear-675           [-1, 16, 16, 96]          36,960
     FeedForward-676           [-1, 16, 16, 96]               0
       LayerNorm-677           [-1, 16, 16, 96]             192
         PreNorm-678           [-1, 16, 16, 96]               0
        Residual-679           [-1, 16, 16, 96]               0
       SwinBlock-680           [-1, 16, 16, 96]               0
          Linear-681          [-1, 16, 16, 576]          55,296
          Linear-682           [-1, 16, 16, 96]          18,528
 WindowAttention-683           [-1, 16, 16, 96]               0
       LayerNorm-684           [-1, 16, 16, 96]             192
         PreNorm-685           [-1, 16, 16, 96]               0
        Residual-686           [-1, 16, 16, 96]               0
          Linear-687          [-1, 16, 16, 384]          37,248
            GELU-688          [-1, 16, 16, 384]               0
          Linear-689           [-1, 16, 16, 96]          36,960
     FeedForward-690           [-1, 16, 16, 96]               0
       LayerNorm-691           [-1, 16, 16, 96]             192
```

```
         PreNorm-692           [-1, 16, 16, 96]               0
        Residual-693           [-1, 16, 16, 96]               0
       SwinBlock-694           [-1, 16, 16, 96]               0
     cyclicShift-695           [-1, 16, 16, 96]               0
          Linear-696          [-1, 16, 16, 576]          55,296
          Linear-697           [-1, 16, 16, 96]          18,528
     cyclicShift-698           [-1, 16, 16, 96]               0
 WindowAttention-699           [-1, 16, 16, 96]               0
       LayerNorm-700           [-1, 16, 16, 96]             192
         PreNorm-701           [-1, 16, 16, 96]               0
        Residual-702           [-1, 16, 16, 96]               0
          Linear-703          [-1, 16, 16, 384]          37,248
            GELU-704          [-1, 16, 16, 384]               0
          Linear-705           [-1, 16, 16, 96]          36,960
     FeedForward-706           [-1, 16, 16, 96]               0
       LayerNorm-707           [-1, 16, 16, 96]             192
         PreNorm-708           [-1, 16, 16, 96]               0
        Residual-709           [-1, 16, 16, 96]               0
       SwinBlock-710           [-1, 16, 16, 96]               0
     StageModule-711           [-1, 96, 16, 16]               0
 ConvTranspose2d-712           [-1, 96, 64, 64]         295,008
  PatchExpanding-713           [-1, 64, 64, 96]               0
          Linear-714          [-1, 64, 64, 288]          27,648
          Linear-715           [-1, 64, 64, 96]           9,312
 WindowAttention-716           [-1, 64, 64, 96]               0
       LayerNorm-717           [-1, 64, 64, 96]             192
         PreNorm-718           [-1, 64, 64, 96]               0
        Residual-719           [-1, 64, 64, 96]               0
          Linear-720          [-1, 64, 64, 384]          37,248
            GELU-721          [-1, 64, 64, 384]               0
          Linear-722           [-1, 64, 64, 96]          36,960
     FeedForward-723           [-1, 64, 64, 96]               0
       LayerNorm-724           [-1, 64, 64, 96]             192
         PreNorm-725           [-1, 64, 64, 96]               0
        Residual-726           [-1, 64, 64, 96]               0
       SwinBlock-727           [-1, 64, 64, 96]               0
     cyclicShift-728           [-1, 64, 64, 96]               0
          Linear-729          [-1, 64, 64, 288]          27,648
          Linear-730           [-1, 64, 64, 96]           9,312
     cyclicShift-731           [-1, 64, 64, 96]               0
 WindowAttention-732           [-1, 64, 64, 96]               0
       LayerNorm-733           [-1, 64, 64, 96]             192
         PreNorm-734           [-1, 64, 64, 96]               0
        Residual-735           [-1, 64, 64, 96]               0
          Linear-736          [-1, 64, 64, 384]          37,248
            GELU-737          [-1, 64, 64, 384]               0
          Linear-738           [-1, 64, 64, 96]          36,960
     FeedForward-739           [-1, 64, 64, 96]               0
       LayerNorm-740           [-1, 64, 64, 96]             192
         PreNorm-741           [-1, 64, 64, 96]               0
        Residual-742           [-1, 64, 64, 96]               0
       SwinBlock-743           [-1, 64, 64, 96]               0
     StageModule-744           [-1, 96, 64, 64]               0
================================================================
Total params: 62,169,888
Trainable params: 62,169,888
Non-trainable params: 0
----------------------------------------------------------------
```

```
Input size (MB): 0.05
Forward/backward pass size (MB): 277.52
Params size (MB): 237.16
Estimated Total Size (MB): 514.73
----------------------------------------------------------------
```