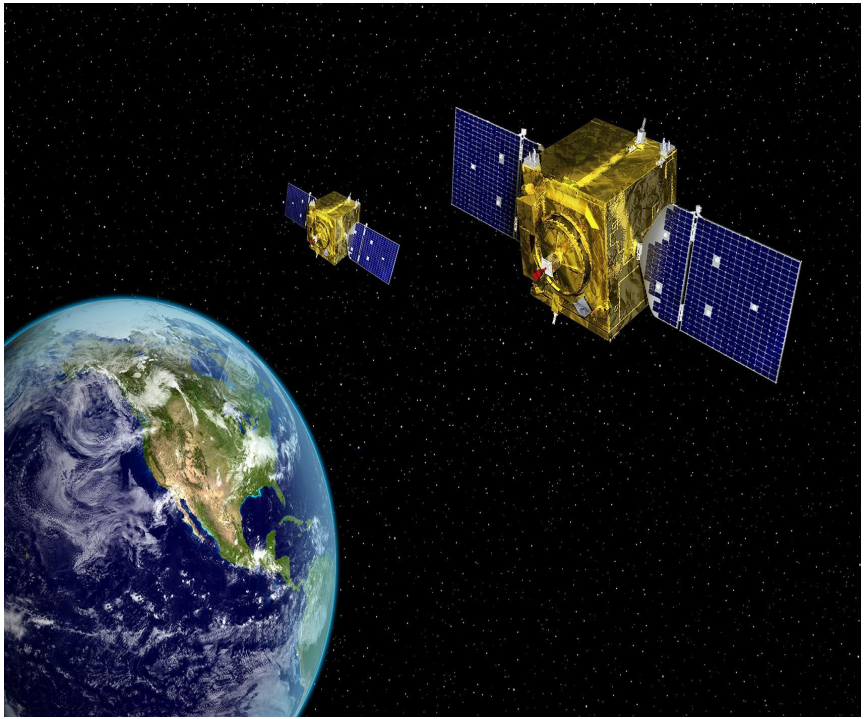
Semantic Segmentation of Satellite Images with Deep Learning Method

Weili Shi, Daiqing Qi
2022-12-02

- Introduction
- Related Work
- Method
- Experiment
- Conslusion

Introduction

- Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance.
- One outcome of the remote sensing is the satellite images, which contains abundant spatial details and rich potential semantic contents for analysis.

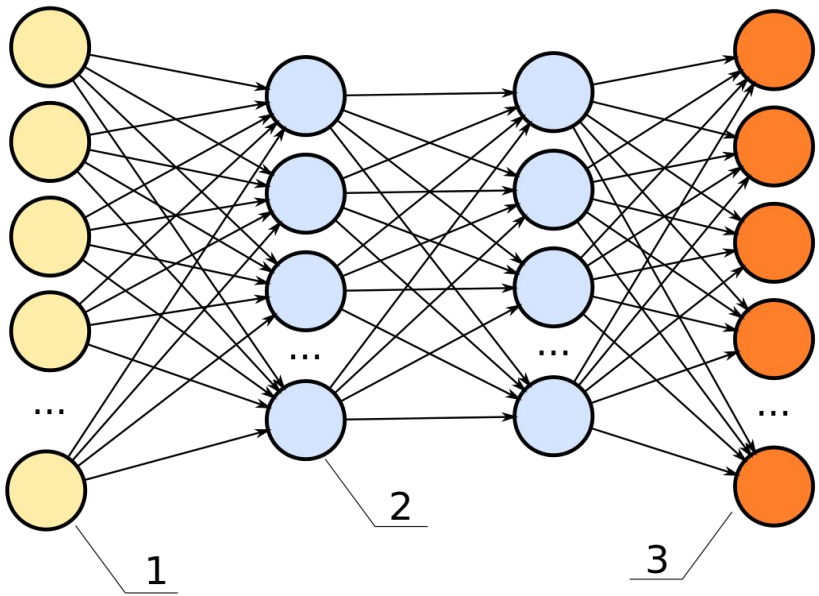


Introduction

- Recently, a growing wave of deep learning technology, has achieved great success in computer vision tasks such as image classification, object detection and semantic segmentation.

- It is tempting to apply the latest deep learning methods on satellite image analysis such as scene detection and segmentation. And it has arouse wide interests among researchers.

Introduction



- Apart from the dominant CNN model, some recent work started to adopt transformer-based method to tackle with semantic segmentation task on satellite images.
- Transformer-based method has shown promising results due to the distinct representative ability.

Introduction

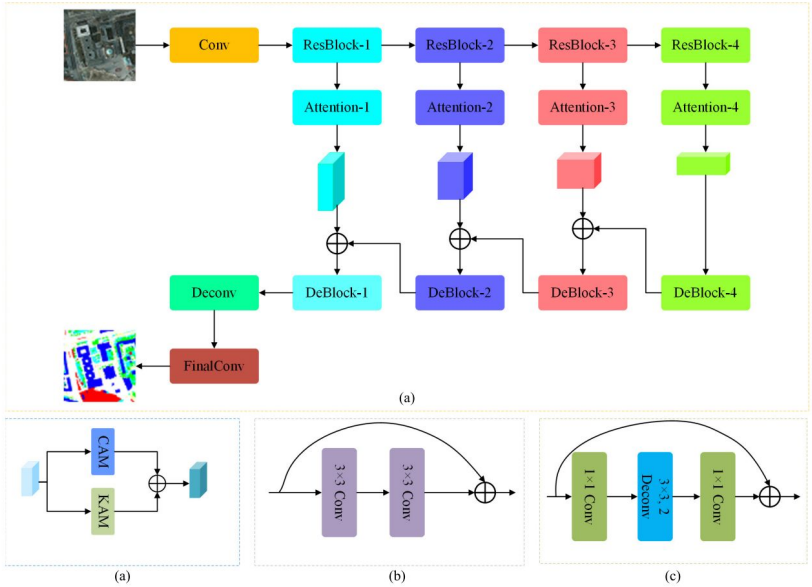
- A typical segmentation framework consists of an encoder and a decoder.

- In our project, We investigate how different encoder would affect the performance of the model. We compare CNN-based encoder and transformer-based encoder.

- This problem is nontrivial since different encoders may affect the final performance of the framework.

Related work

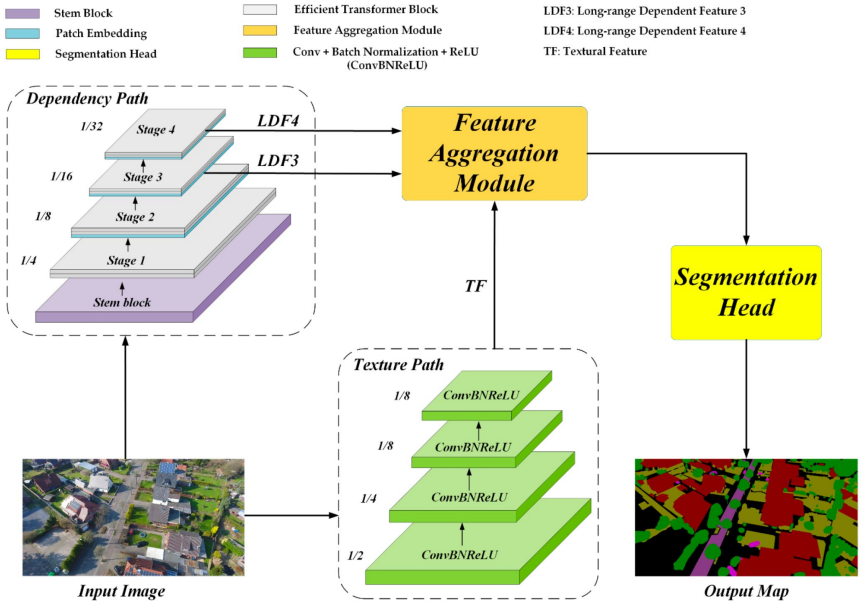
MultiAttention network (MANet) extracts contextual dependencies through multiple efficient attention modules. A novel attention mechanism of kernel attention with linear complexity is proposed to alleviate the large computational demand in attention.



MANet, Li et al.

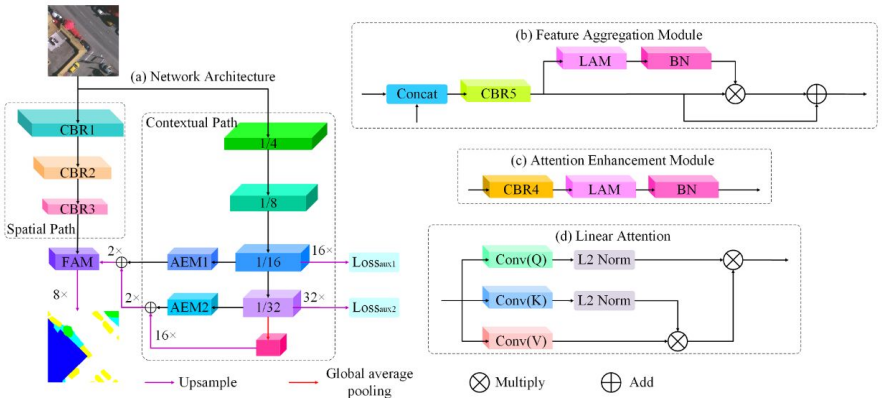
Related work

Bilateral Awareness Network contains a dependency path and a texture path to fully capture the long-range relationships and fine-grained details in VFR images. Specifically, the dependency path is conducted based on the ResT, a novel Transformer backbone with memory-efficient multi-head self-attention, while the texture path is built on the stacked convolution operation.

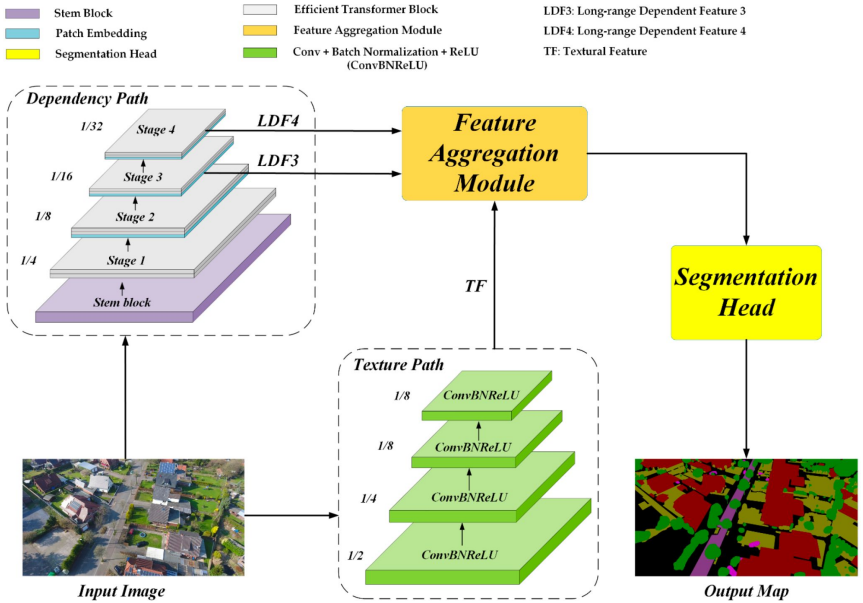


BANet, Wang et al.

Related work



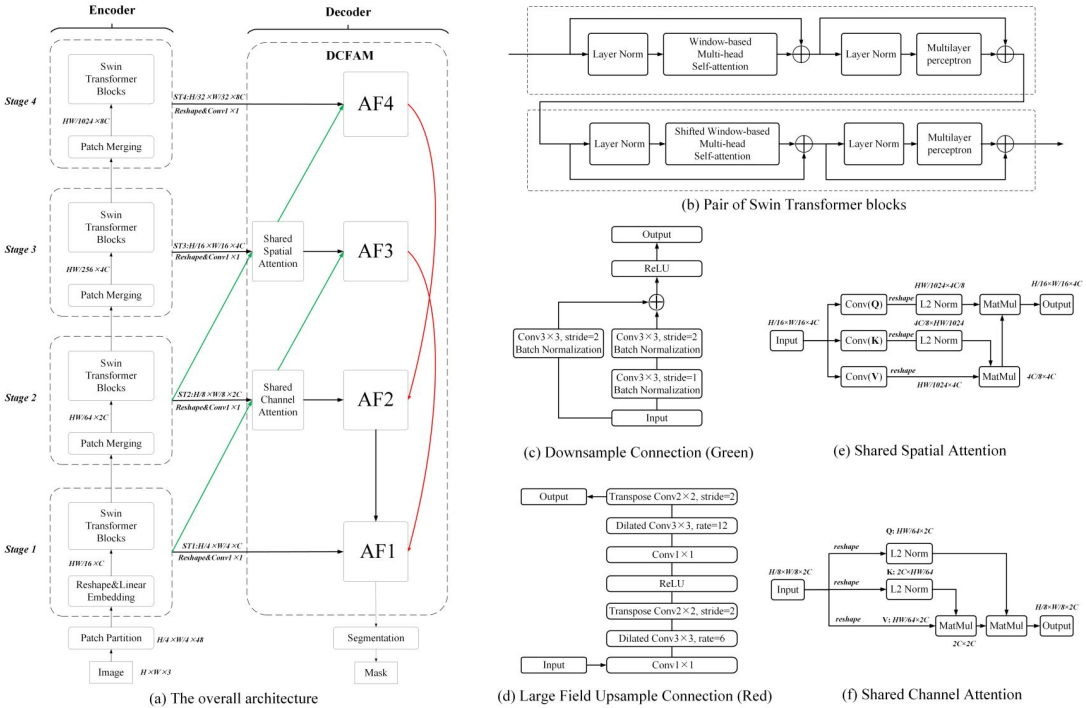
ABCNet, Li et al.



BANet, Wang et al.

Related work

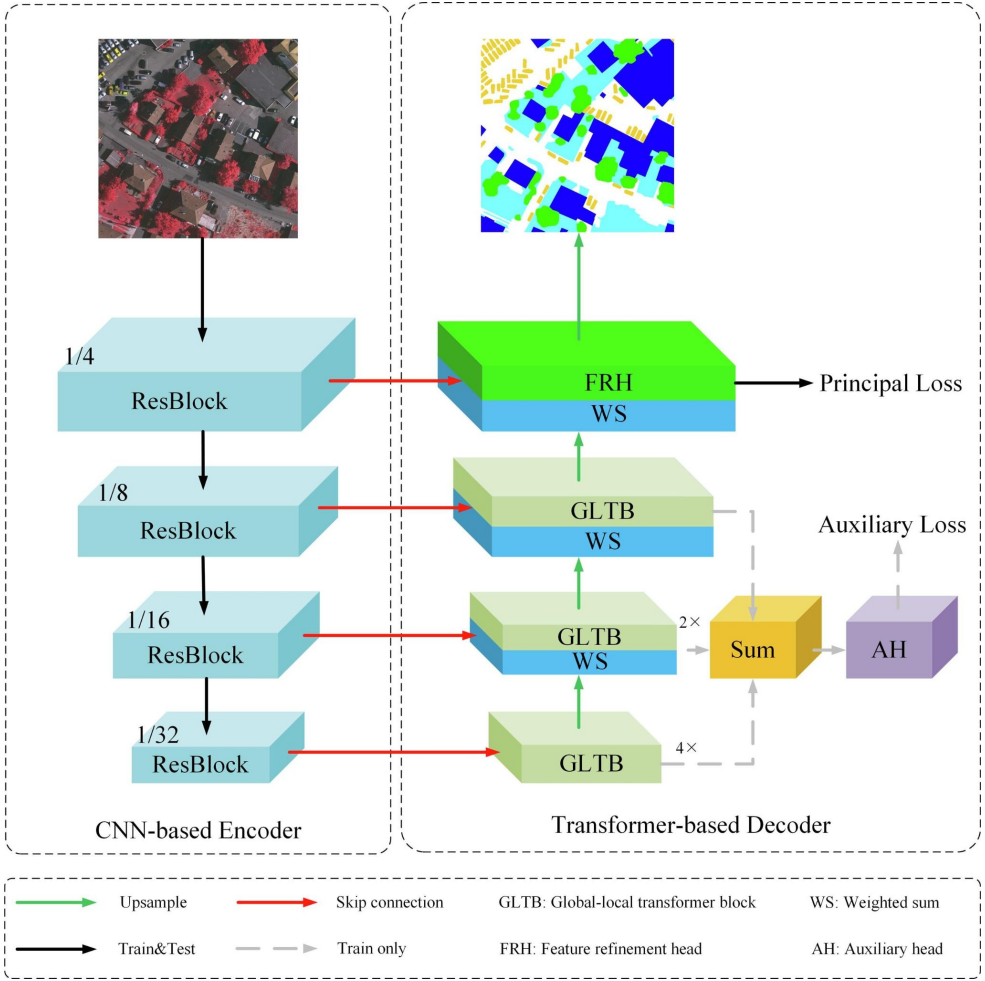
DC-Swin introduces the Swin Transformer as the backbone to extract the context information and design a novel decoder of densely connected feature aggregation module (DCFAM) to restore the resolution and produce the segmentation map.



DC-Swin, Wang et al.

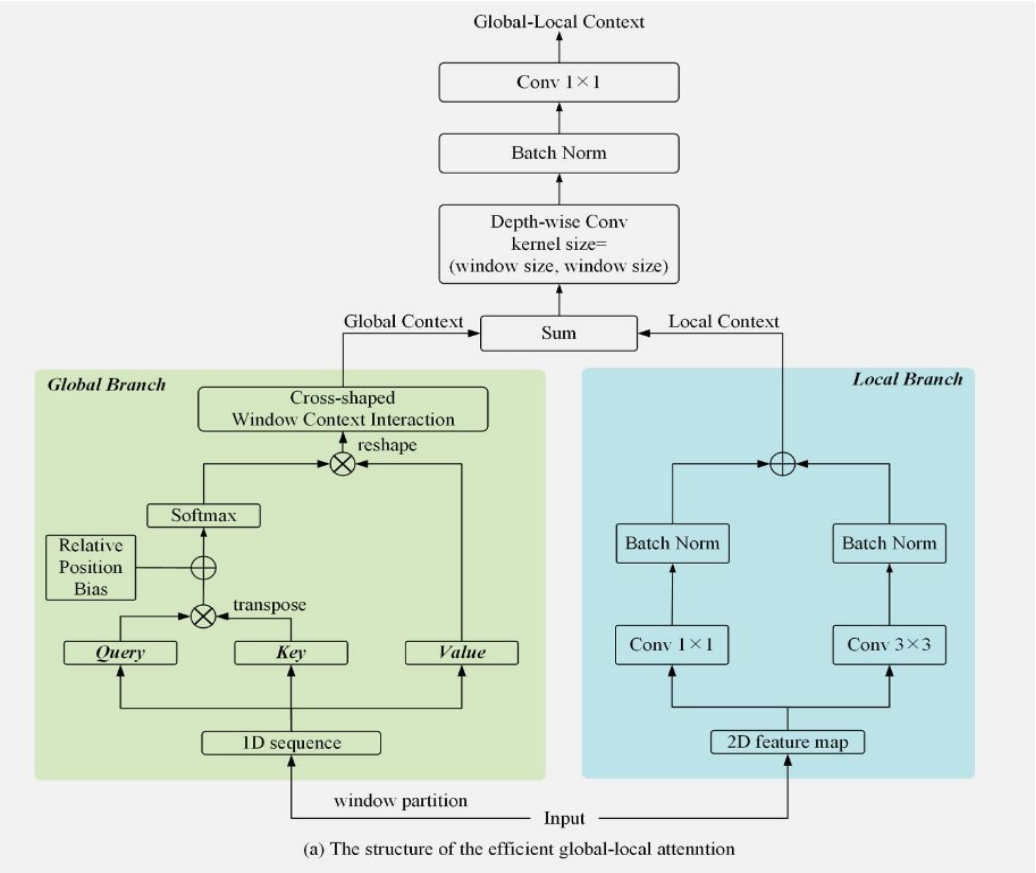
Method

- In our project, we adopt Unetformer as our framework. Unetformer is a UNet-like framework. It consists of a CNN-based encoder and transformer-based encoder.



Method

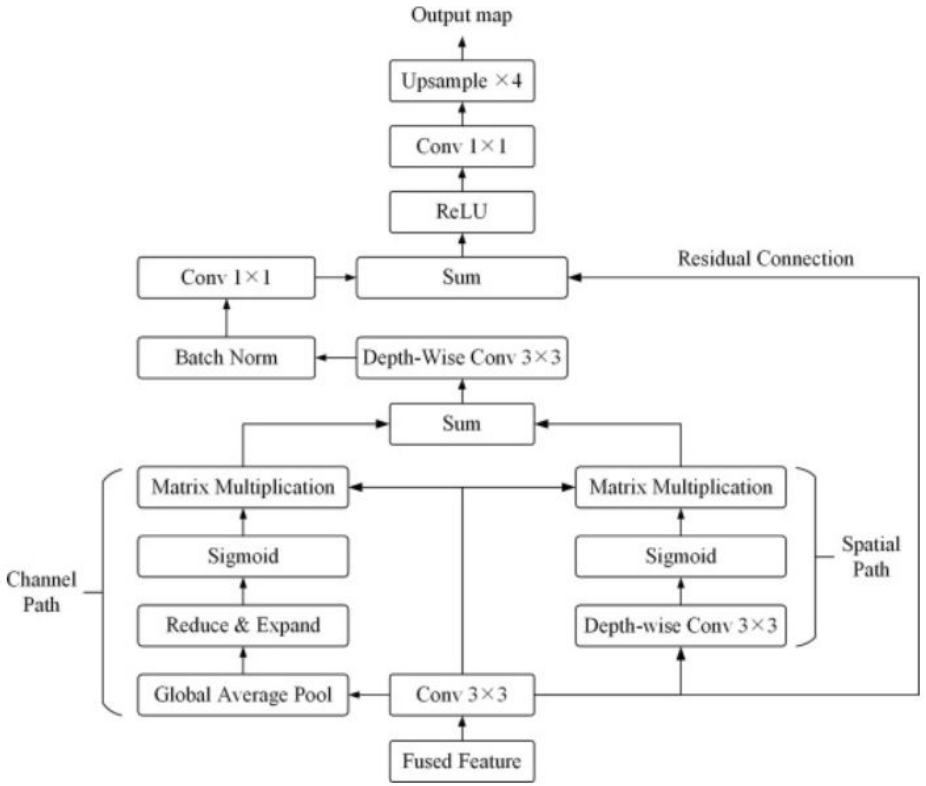
- Illustration of the efficient global–local attention.



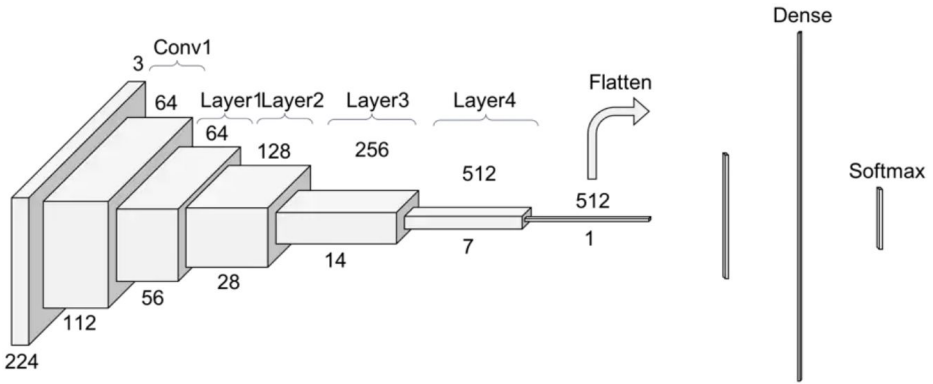
(a) The structure of the efficient global-local attenntion

Method

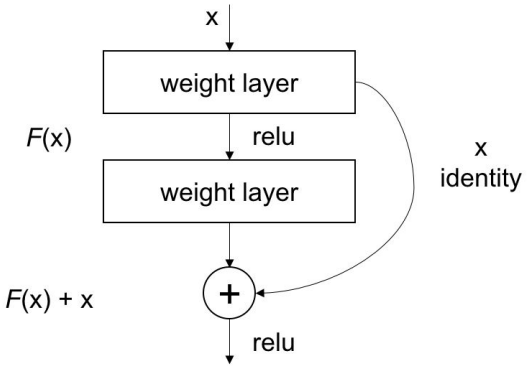
- Illustration of the feature refinement head.



Method



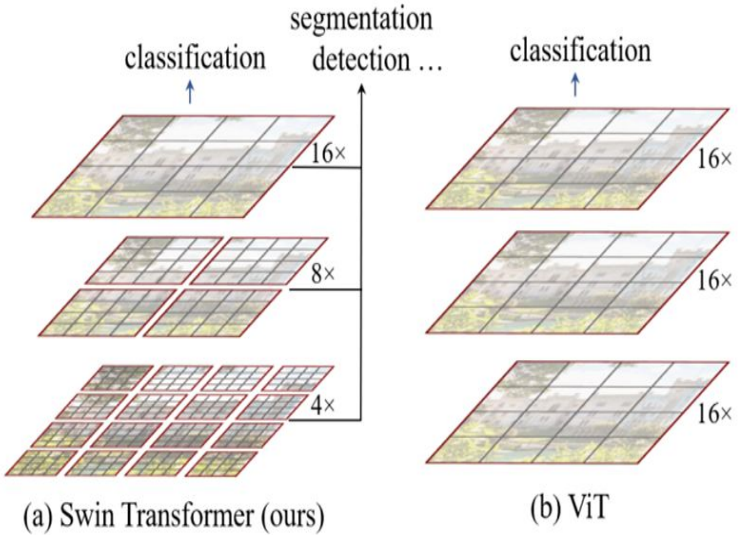
ResNet-34

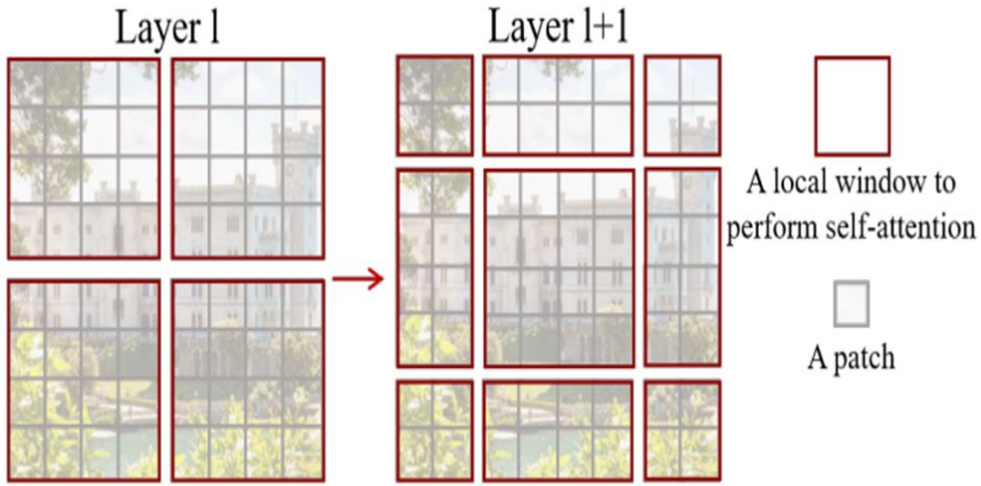


Residual blocks

Method

- We replace ResNet in the Unerformer with swin transformer.





Method

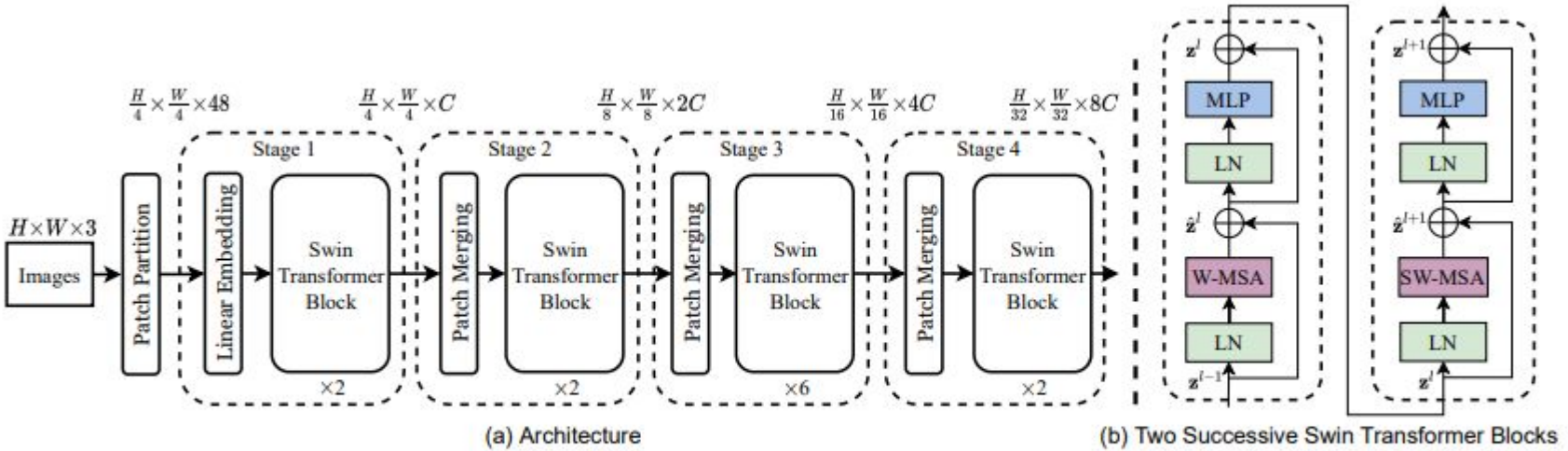


Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.

Experiment

- Dataset: Vaihingen and Potsdam

- Potsdam: 62 for training and 14 for validation
- Vaihingen : 25 for training and 8 for validation

Experiment

- The size of the original image is too large (6000 x 6000)。It should be split into smalle-sized images (1024 x1024).

Experiment

- We use A100 GPU for training.
- The metrics includes mIoU (mean intersection of union), mean F1 score and OA (overall accuracy)

Experiment

Potsdam

encoder	mIoU	F1	OA
CNN	83.3	90.6	90.5
Transformer	81.4	89.5	88.9

Vaihingen

encoder	mIoU	F1	OA
CNN	85.1	91.8	92.6
Transformer	84.5	91.5	92.2

Experiment

original image

ground-truth mask

Experiment

CNN

Transformer

Conclusion

- We investigate the performance of CNN-based and transformer-based encoder on the segmentation of the satellite images. The results show that the two encoders are almost equivalent and the difference of them are trivial in terms of the performance

Thank you