



**ĐẠI HỌC QUỐC GIA TP. HCM  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**ĐỒ ÁN**

**CHUYÊN ĐỀ NGHIÊN CỨU VÀ ỨNG DỤNG VỀ THỊ GIÁC MÁY TÍNH**

# **ỨNG DỤNG SWIN TRANSFORMER TRONG PHÂN LOẠI VẤN ĐỀ CỦA CHẤT LƯỢNG HÌNH ẢNH**

**Giảng viên bộ môn: TS. Mai Tiến Dũng**

**Học viên: Đinh Văn Hoàn**

**Vũ Bảo Quốc**

01

Introduction

02

Method

03

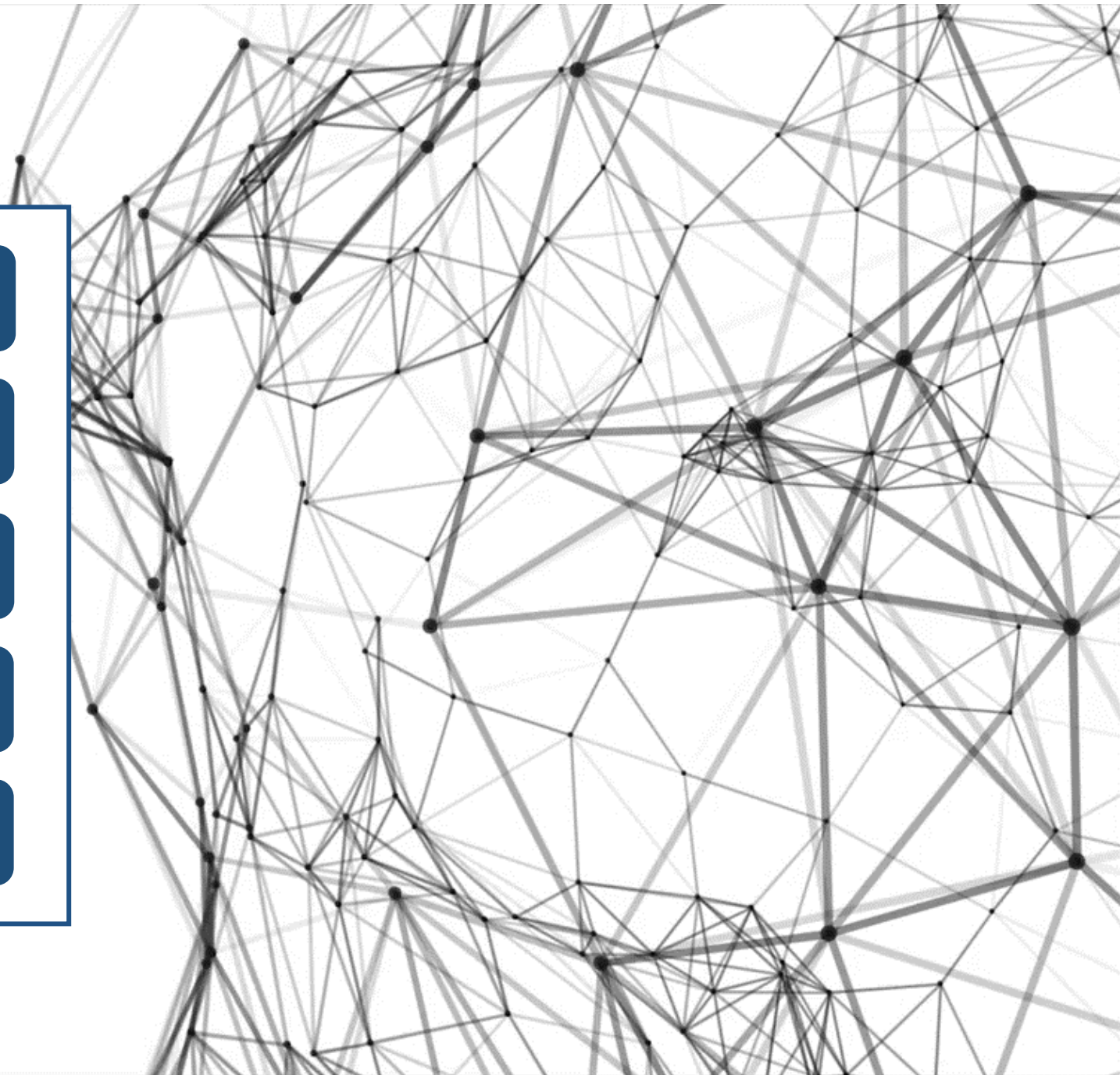
Experiment

04

Conclusion

05

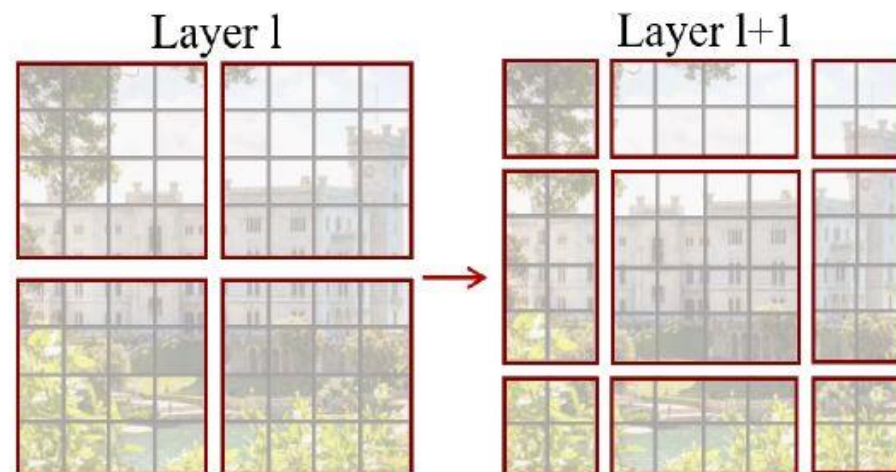
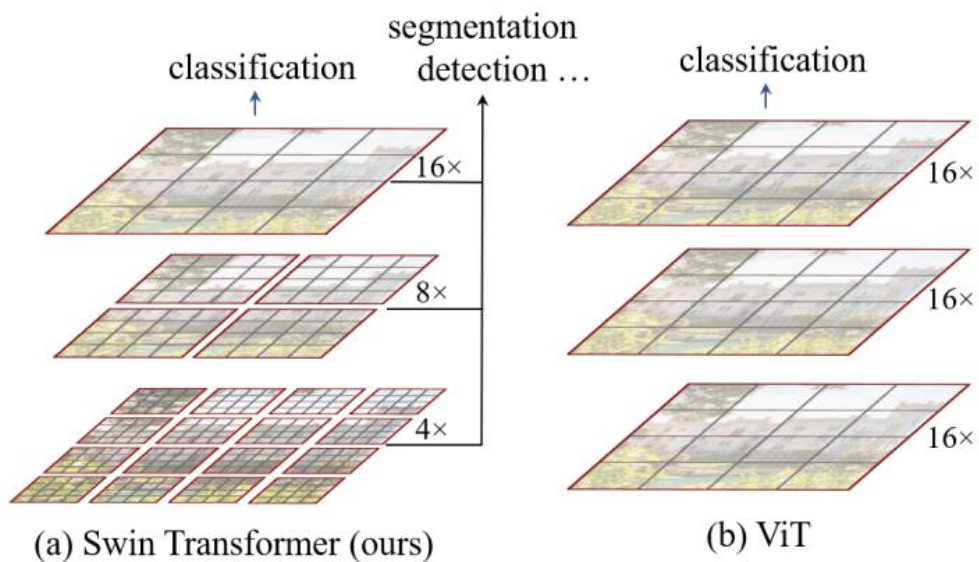
Q&amp;A



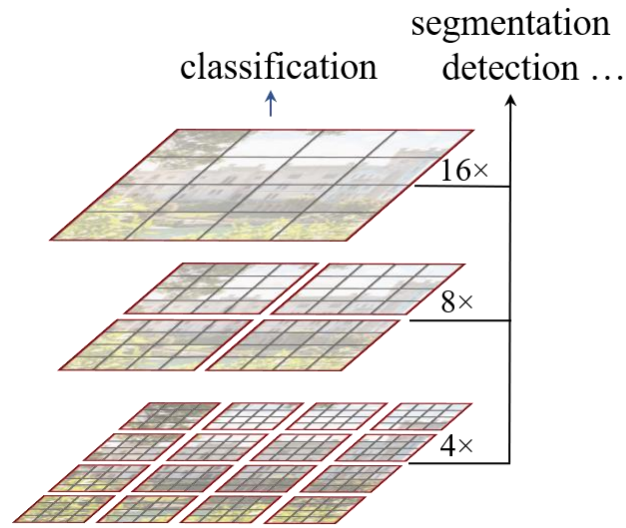


- Optimize Transformer to achieve the better performance.
- Patch problem: different patch partition between two layers -- cyclic shift
- Have the ability to replace Convolution

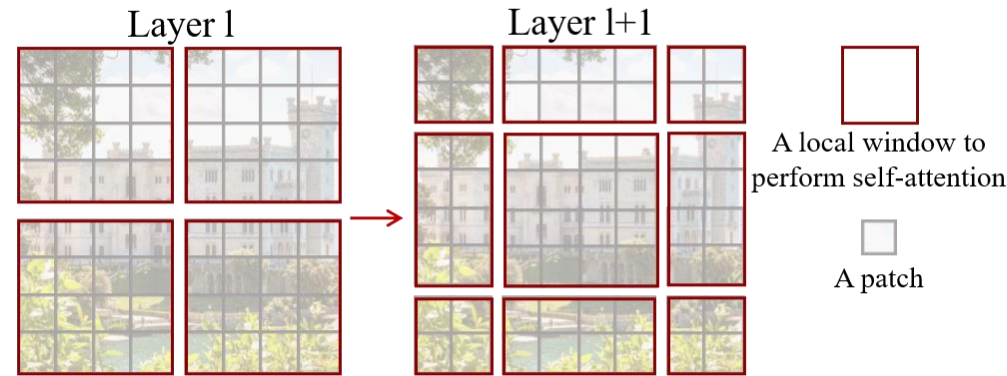
	ImageNet		COCO		ADE20k
	top-1	top-5	AP <sup>box</sup>	AP <sup>mask</sup>	mIoU
w/o shifting	80.2	95.1	47.7	41.5	43.3
shifted windows	<b>81.3</b>	<b>95.6</b>	<b>50.5</b>	<b>43.7</b>	<b>46.1</b>
no pos.	80.1	94.9	49.2	42.6	43.8
abs. pos.	80.5	95.2	49.0	42.4	43.2
abs.+rel. pos.	81.3	95.6	50.2	43.4	44.0
rel. pos. w/o app.	79.3	94.7	48.2	41.9	44.1
rel. pos.	<b>81.3</b>	<b>95.6</b>	<b>50.5</b>	<b>43.7</b>	<b>46.1</b>



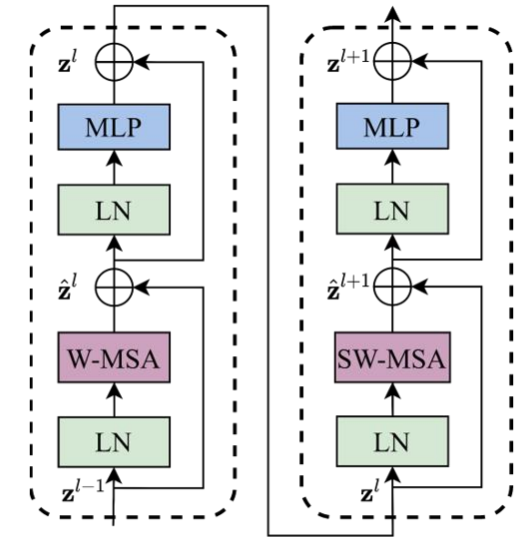
# Swin Transformer (Shifted Window Transformer)



(a) Swin Transformer



(b) Shifted Window



(c) Two Successive Swin Transformer Blocks

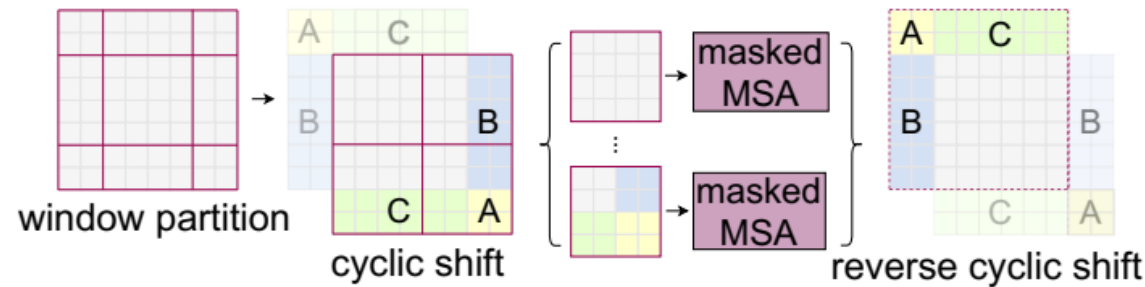
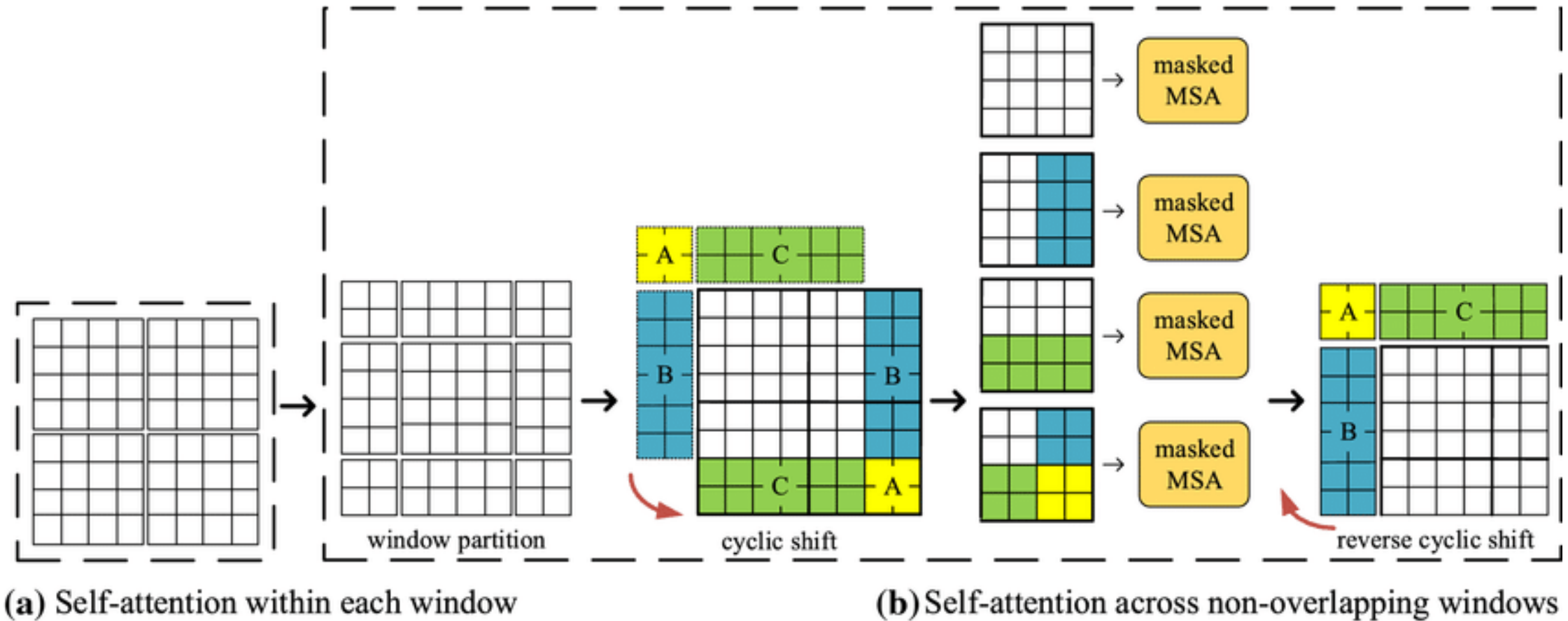


Figure 4. Illustration of an efficient batch computation approach for self-attention in shifted window partitioning.

# Swin Transformer (Shifted Window Transformer)



# Overall Architecture

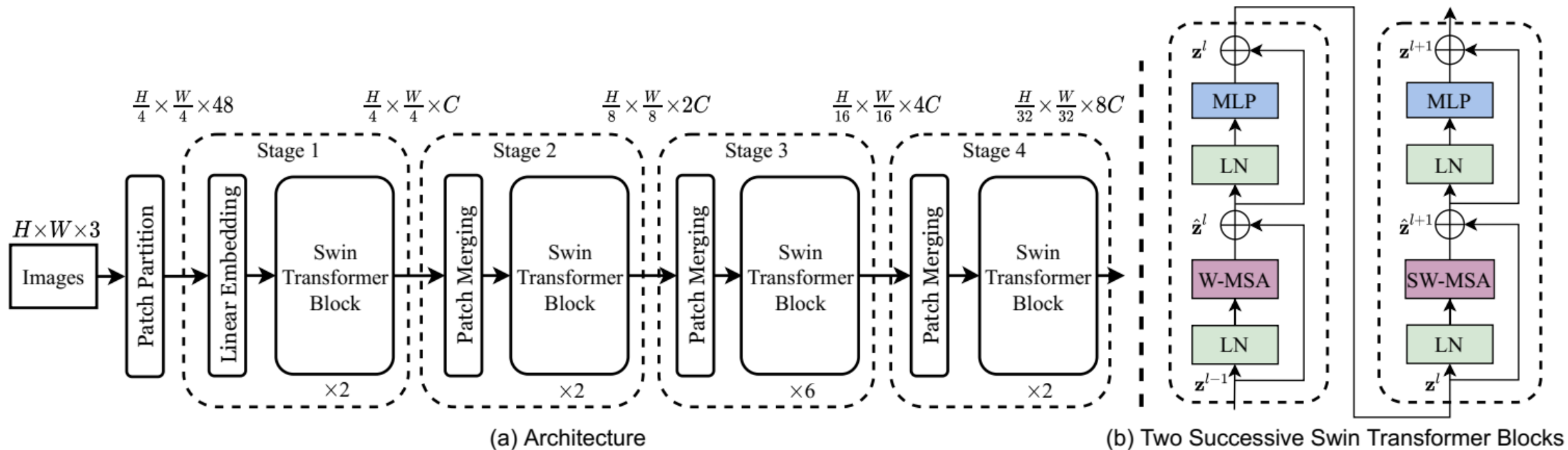
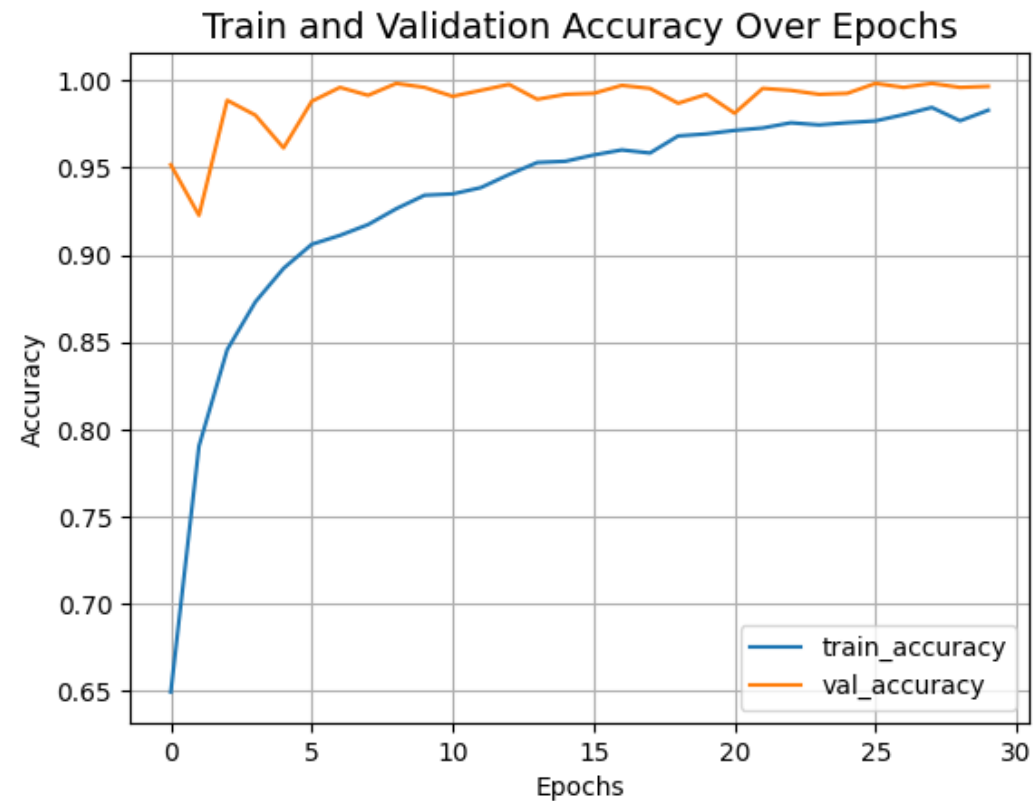
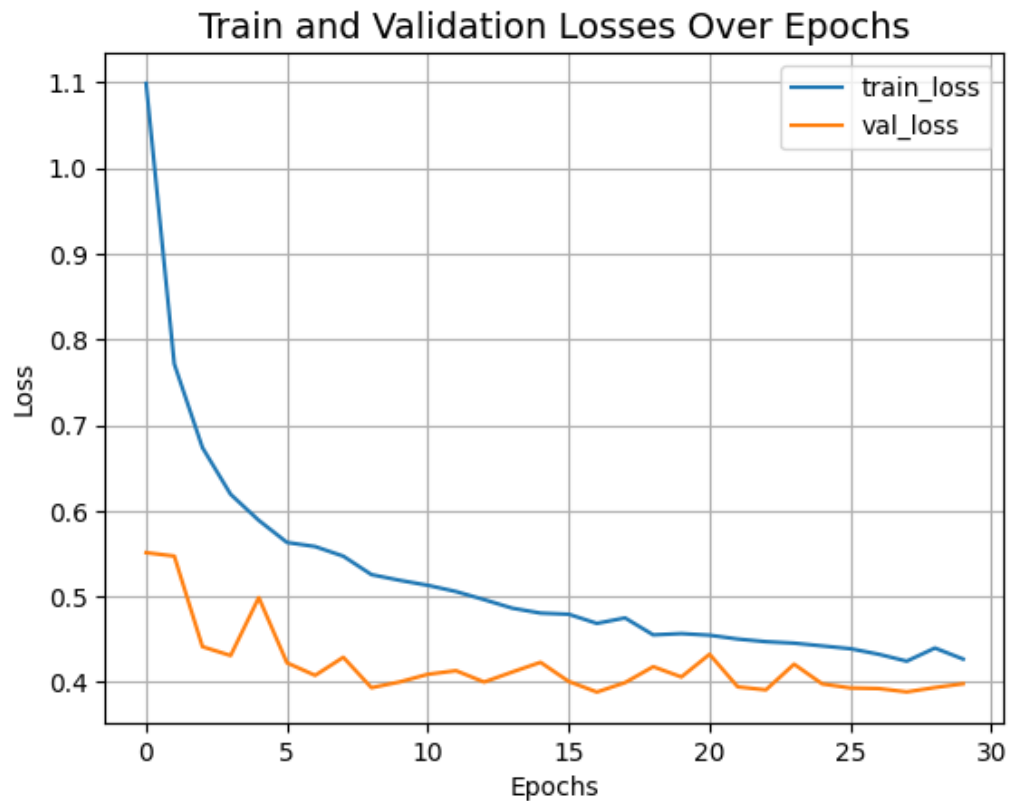


Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.

## Dataset

Dataset	Số lượng ảnh (train/test)	Nguồn thu thập
Defocused-blurred	734 (734/315)	DPDD (Dual-Pixel Defocus Deblurring)
Motion-blurred	5.199 (5.039/2.160)	GoPro, HIDE, RealBlurR, RealBlurJ_test
Noise images	3,160 (2.212/948)	SIDD (Smartphone-image-denoising-dataset)
Rain images	13,711 (9.597/4.114)	Synthetic rain datasets (tập Rain13K)

## Validation



## Evaluation

236/236 [=====] - 17s 73ms/step - loss: 0.4814 - accuracy: 0.9549 - top-5-accuracy: 1.0000

Test loss: 0.48

Test accuracy: 95.49%

Test top 5 accuracy: 100.0%





1/1 [=====] - 0s 27ms/step  
Predicted Label: Defocused blur



1/1 [=====] - 0s 27ms/step  
Predicted Label: Noise



1/1 [=====] - 0s 27ms/step  
Predicted Label: Rain



1/1 [=====] - 0s 26ms/step  
Predicted Label: Rain

**Hình 3.2. Kết quả thử nghiệm ngẫu nhiên trên 4 ảnh được thu thập ngẫu nhiên trên internet.**

- Mô hình Swin Transformer đạt hiệu suất cao trong quá trình huấn luyện, với độ chính xác và độ tin cậy tăng đáng kể qua các epoch.
- Khả năng học tốt của Swin Transformer trên dữ liệu chất lượng hình ảnh được thể hiện rõ.
- Kết quả kiểm thử cho thấy mô hình có khả năng tổng quát hóa tốt trên dữ liệu mới, với độ chính xác ổn định và loss thấp.
- Mặc dù hiệu suất cao, mô hình vẫn mắc phải một số sai sót trong việc phân loại ảnh từ dữ liệu ngẫu nhiên thu thập từ internet.

## REFERENCES

- Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022). Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5728-5739).
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).
- Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., & Shao, L. (2022). Learning enriched features for fast image restoration and enhancement. IEEE transactions on pattern analysis and machine intelligence, 45(2), 1934-1948.
- Cai, Y., Hu, X., Wang, H., Zhang, Y., Pfister, H., & Wei, D. (2021). Learning to generate realistic noisy images via pixel-level noise-aware adversarial training. Advances in Neural Information Processing Systems, 34, 3259-3270.
- Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., & Li, H. (2022). Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 17683-17693).
- Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., ... & Gao, W. (2021). Pre-trained image processing transformer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12299-12310).

