

# Auto White Balance for Multi-illuminant Scene

팀 JYP

팀원: 박윤정, 유수민, 조운수

지도교수: 김선주

조교: 김동영

# 목차

- 연구 주제 및 필요성
- U-Net, U-Net3+
- SWIN Transformer + UperNet
- 실험 결과
- 진행 시 문제점 및 보완 계획
- 역할 배분

# 연구 주제

- LSMI dataset을 사용해 개선된 AWB 모델 탐색
- 기존 U-Net 모델을 수정해 U-Net3+ 모델 적용
- Vision Transformer를 활용한 심화 모델 구현

# 연구의 필요성

- 기존 연구의 한계
  - Multi-illuminants 장면에 대한 연구 부족
- Computer Vision 연구에 사용되는 이미지 개선
  - AWB가 적용된 이미지는 딥러닝 모델의 학습을 개선시킴

# 기존 연구

LSMI dataset 연구에서 사용한 U-net 모델의 경우

- 이미지의 illuminants의 chrominance( $u, v$ ) 예측
- Pixel별 예측

# 이번 연구

- U-net (baseline 모델)
- U-net3+
- SWIN Transformer + UperNet

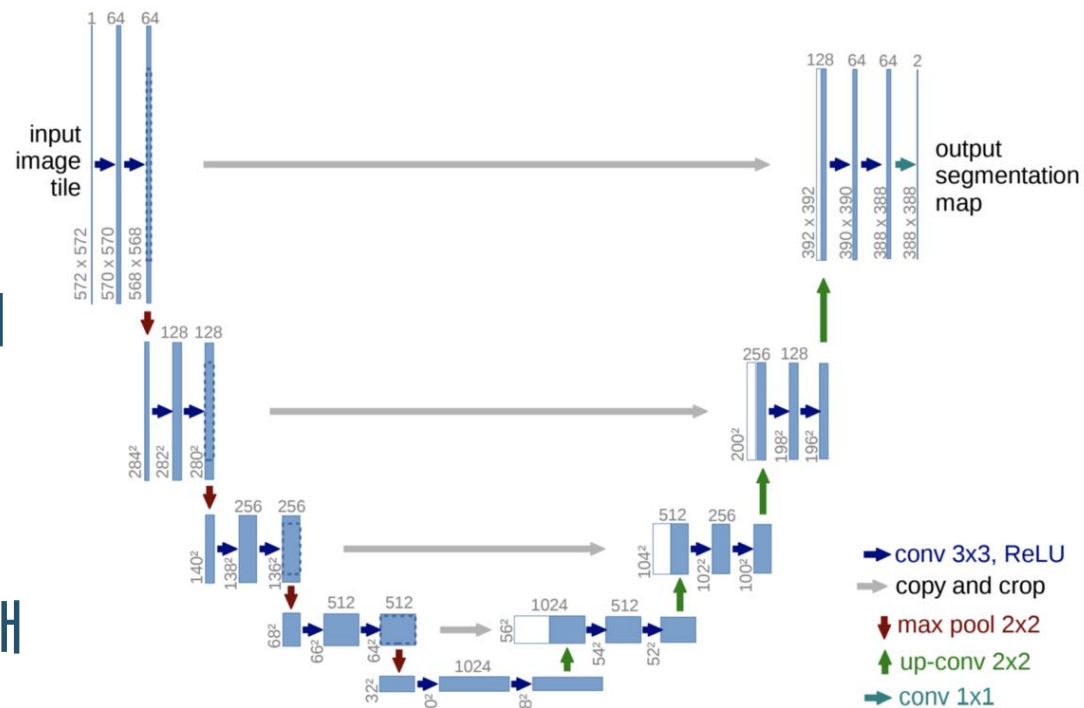
→ Semantic Segmentation task 수행

# U-Net

▷ Biomedical 분야에서 Image Segmentation을 목적으로 제안됨

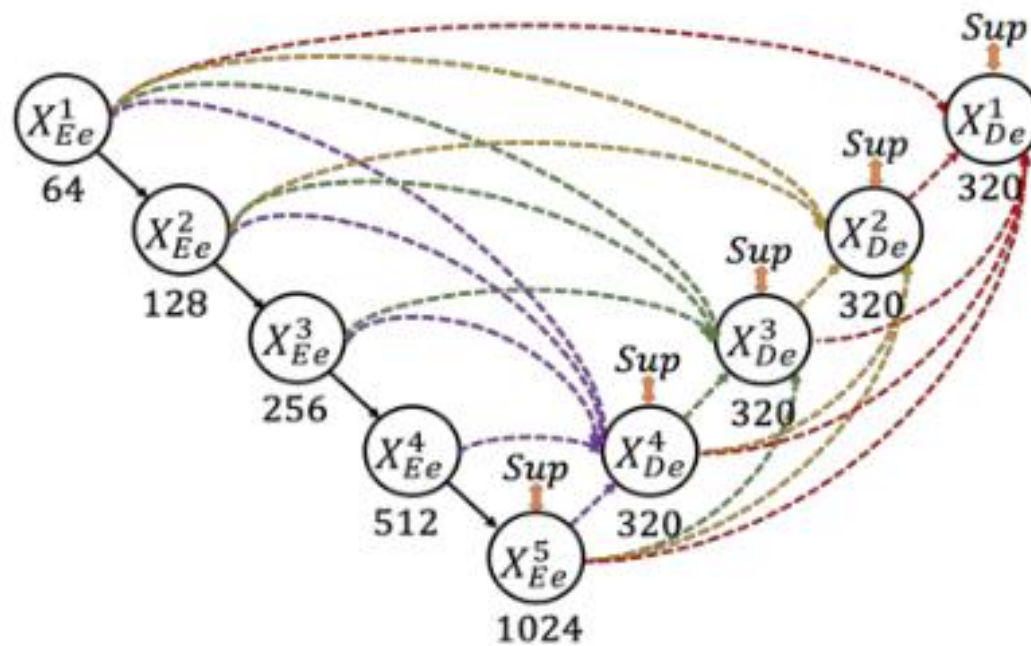
(1) Contracting path: 입력 이미지의 컨텍스트(이웃 픽셀간 정보와 이미지 문맥)를 파악

(2) Expanding path: Feature map을 Upsampling하고 컨텍스트와 결합해 정확한 localization을 함



# U-Net3+

- 의료 영상을 다루기 위해 개발된 encoder-decoder 구조의 모델
- Image segmentation에 있어 우수, pixel-level task에서 널리 사용됨.



(c) UNet 3+

(Full-scale skip connections)



# U-Net, U-Net3+ 학습, 테스트 결과

2000 epoch 후 test 성능 비교 (MAE\_illum mean)

	모델	Batch size	Learning rate	Learning rate decay	Initial weight	MAE_illum mean
1	U-net	32	0.0005	1200	파이토치 기본	2.0421
2	U-net3+	8	0.0005	1200	파이토치 기본	2.4012
3	U-net3+	8	0.00055	1200	파이토치 기본	2.4102
4	U-net3+	8	0.0005	1200	Kaiming	2.4187
5	U-net3+	8	0.0005	800	Kaiming	2.4602
6	U-net	8	0.0005	1200	Kaiming	3.4980

# SWIN transformer + UperNet

## Encoder - Decoder 구조 사용

1. backbone: SWIN transformer  
- feature 추출
2. decode\_head: UperNet  
- upsampling

### 4.3. Semantic Segmentation on ADE20K

ADE20K		val	test	#param.	FLOPs	FPS
Method	Backbone	mIoU	score			
DANet [23]	ResNet-101	45.2	-	69M	1119G	15.2
DLab.v3+ [11]	ResNet-101	44.1	-	63M	1021G	16.0
ACNet [24]	ResNet-101	45.9	38.5	-	-	-
DNL [71]	ResNet-101	46.0	56.2	69M	1249G	14.8
OCRNet [73]	ResNet-101	45.3	56.0	56M	923G	19.3
UperNet [69]	ResNet-101	44.9	-	86M	1029G	20.1
OCRNet [73]	HRNet-w48	45.7	-	71M	664G	12.5
DLab.v3+ [11]	ResNeSt-101	46.9	55.1	66M	1051G	11.9
DLab.v3+ [11]	ResNeSt-200	48.4	-	88M	1381G	8.1
SETR [81]	T-Large <sup>‡</sup>	50.3	61.7	308M	-	-
UperNet	DeiT-S <sup>†</sup>	44.0	-	52M	1099G	16.2
UperNet	Swin-T	46.1	-	60M	945G	18.5
UperNet	Swin-S	49.3	-	81M	1038G	15.2
UperNet	Swin-B <sup>‡</sup>	51.6	-	121M	1841G	8.7
UperNet	Swin-L <sup>‡</sup>	<b>53.5</b>	<b>62.8</b>	234M	3230G	6.2

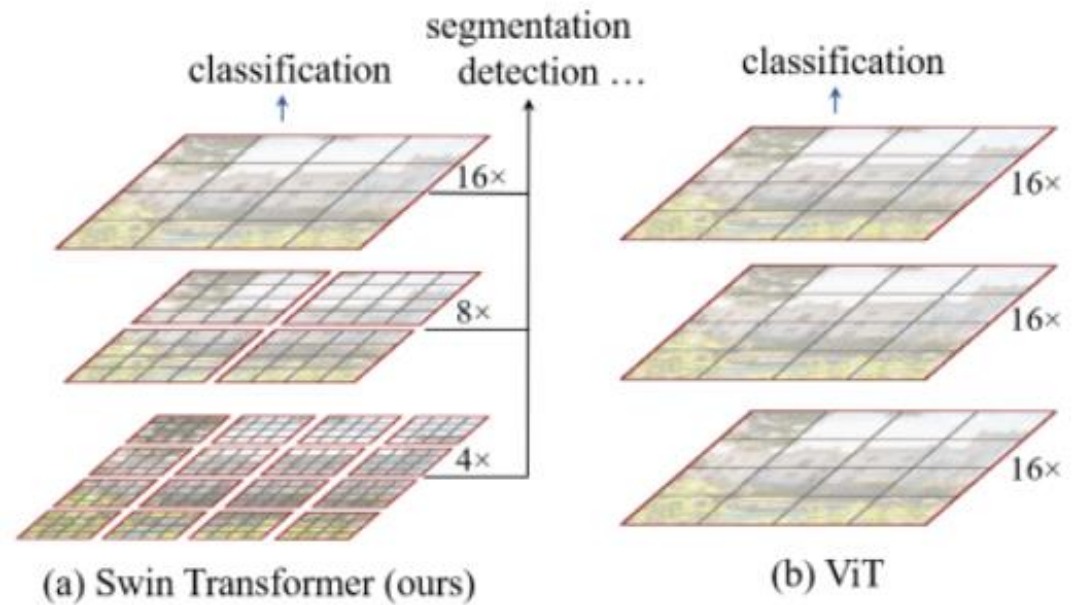
Table 3. Results of semantic segmentation on the ADE20K val and test set. <sup>†</sup> indicates additional deconvolution layers are used to produce hierarchical feature maps. <sup>‡</sup> indicates that the model is pre-trained on ImageNet-22K.

다른 Task(Object Detection, Semantic Segmentation)의 backbone으로 사용했을 때의 성능은 거의 state-of-the-art이다.

# SWIN transformer + UperNet

## SWIN transformer란

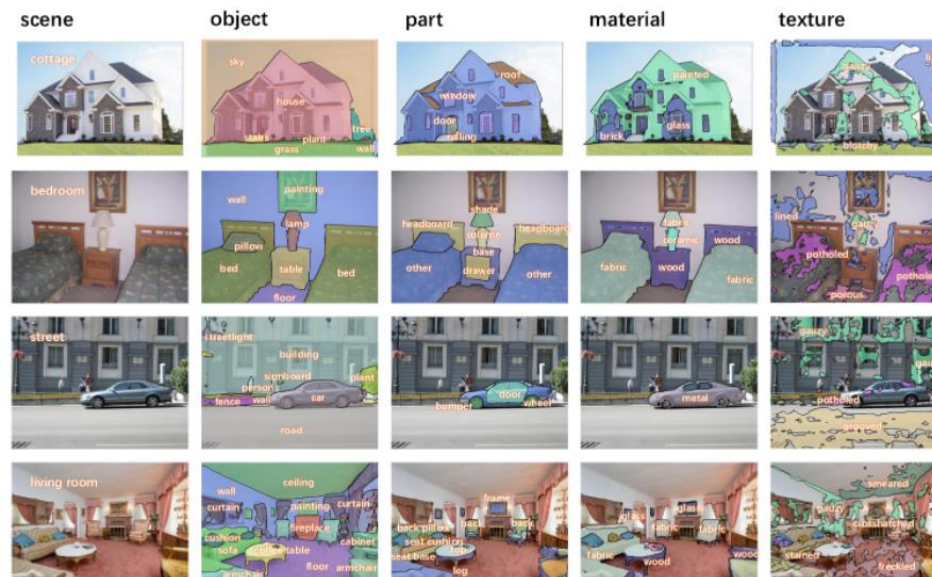
- Vision transformer 중 하나인 Shifted Window Transformer
- ViT보다 계산량이 적고 다양한 scale을 처리 가능
  - 이미지 크기에 선형비례하는 계산량



# SWIN transformer + UperNet

## UperNet이란

- 객체의 세분화된 속성에 따른 segmentation을 동시에 수행하는 모델(multi-task learning)
- Upsampling 과정.  
input image와 크기와 같은  
segmentation map을 만들기 위해 사용



[왼쪽에서 오른쪽으로 (추론 결과) : 장면 분류 및 객체, 부품, 재질 및 텍스처 파싱]






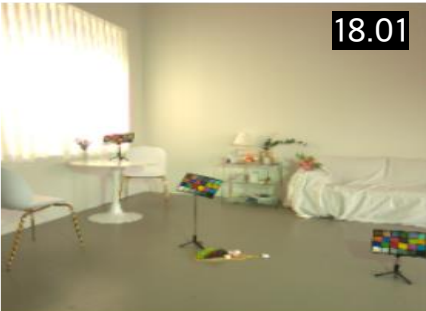
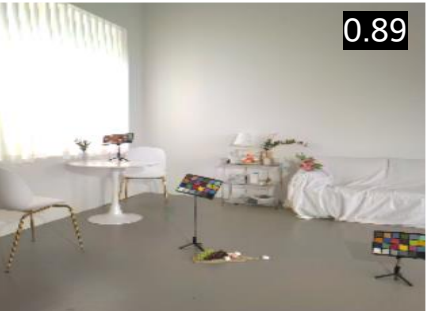
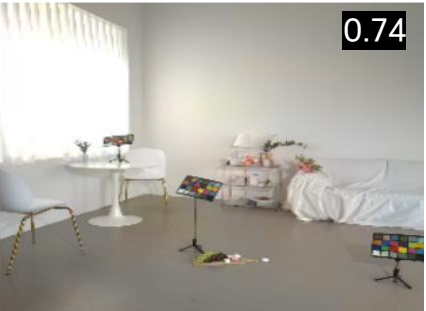
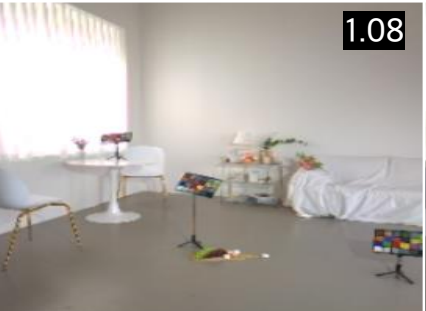
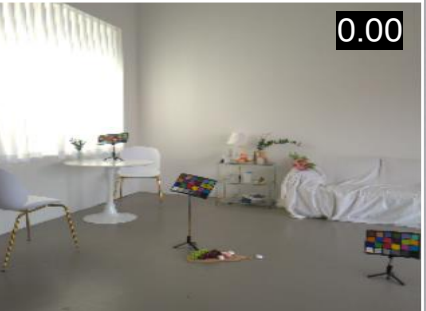




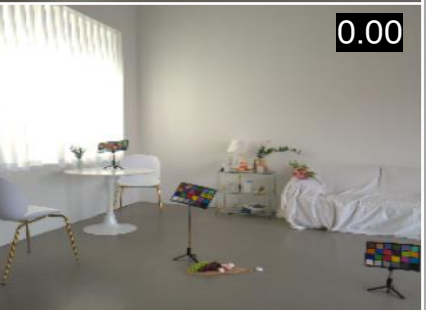
# 실험 결과

## 1. 성능 (MAE)

모델	Min	Mean	Median	Max
U-net	0.80	<b>3.09</b>	<b>2.04</b>	15.19
U-net3+	<b>0.74</b>	3.19	2.40	<b>13.75</b>
SWIN+Upernet	1.03	3.63	2.38	16.91



# 실험 결과

	Input Image	U-Net	U-Net3+	Swin+Upernet	GT
Best MAE illumination for U-Net (1 illum: 1)	 16.84	 0.80	 1.16	 1.08	 0.00
Best MAE illumination for U-net3+ (2 illums: 1,2)	 18.01	 0.89	 0.74	 1.08	 0.00
Best MAE illumination for SWIN+Uper net (1 illum: 1)	 15.32	 1.73	 2.10	 1.03	 0.00

# 실험 결과

	Input Image	U-Net	U-Net3+	Swin+Upernet	GT
Worst MAE illumination for U-Net (2 illums: 1, 3)	 23.54	 15.19	 13.75	 13.44	 0.00
Worst MAE illumination for U-net3+ (2 illums: 1, 3)	 23.54	 15.19	 13.75	 13.44	 0.00
Worst MAE illumination for SWIN+Upernet (1 illum: 1)	 17.25	 10.19	 8.88	 16.91	 0.00




# 실험 결과

	Input Image	U-Net	GT
Low MAE illumination (3 illums: 1, 2, 3)	 17.45	 4.07	 0.00
Mid MAE illumination (2 illums: 1, 3)	 23.39	 7.13	 0.00
High MAE illumination (2 illums: 1, 2)	 22.11	 12.44	 0.00



# 실험 결과

	Input Image	U-Net3+	GT
Low MAE illumination (3 illums: 1, 2, 3)	 19.54	 4.44	 0.00
Mid MAE illumination (2 illums: 1, 2)	 17.01	 7.12	 0.00
High MAE illumination (2 illums: 1, 2)	 22.11	 13.23	 0.00

# 실험 결과

	Input Image	SWIN+Upernet	GT
Low MAE illumination (3 illums: 1, 2, 3)			
Mid MAE illumination (2 illums: 1, 2)			
High MAE illumination (1 illum: 1)			



# 진행 시 문제점

1. U-net, U-net3+ 모델 크기가 24GB
  - Batch size를 8 이하로만 설정 가능
  - 2000 epoch을 도는 데 4일 이상 소모
2. 심화 모델의 성능이 눈에 띄게 향상되지 않음

# 최종 보고서 전까지 수정, 보완할 부분

- SWIN+UperNet모델의 weight initialization과 learning rate 등의 hyperparameter 조정을 통한 모델 개선
- 조명의 개수와 특징에 따른 각 모델의 성능 비교, 분석

# 역할 분배

박윤정

발표 자료 및 자료조사, 개발

유수민

발표 자료 및 자료조사, 개발

조윤수

발표 자료 및 자료조사, 개발

**감사합니다!**