



IMAGE SUPER-RESOLUTION

AI 양재 허브
인공지능 오픈 소스 경진대회

TEAM : 모두연 포레버
박수철. 장진우. 윤성국. 양성모

—

목차

- 대회 개요
- 데이터 특징
- Model Selection
- Methods
- Experiments



대회 개요

대회 개요

■ 주제

> 이미지 초해상화(Image Super-Resolution)를 위한 AI 알고리즘 개발

■ 목적 및 배경

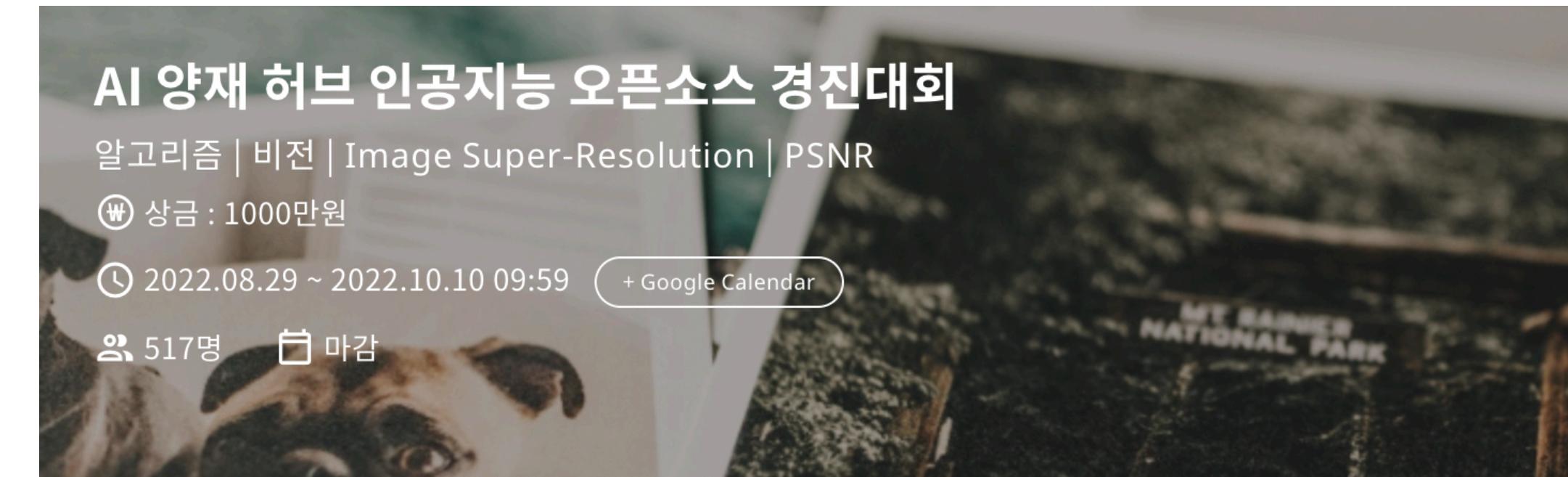
> 인공지능 오픈소스를 활용하여 이미지 초해상화 문제를 해결함으로써 오픈 소스 생태계와 컴퓨터 비전 분야에 기여

■ 설명

> 품질이 저하된 저해상도 촬영 이미지(512x512)를 고품질의 고해상도 촬영 이미지(2048x2048)로 생성

■ 평가 산식

> PSNR(Peak Signal-to-Noise Ratio)

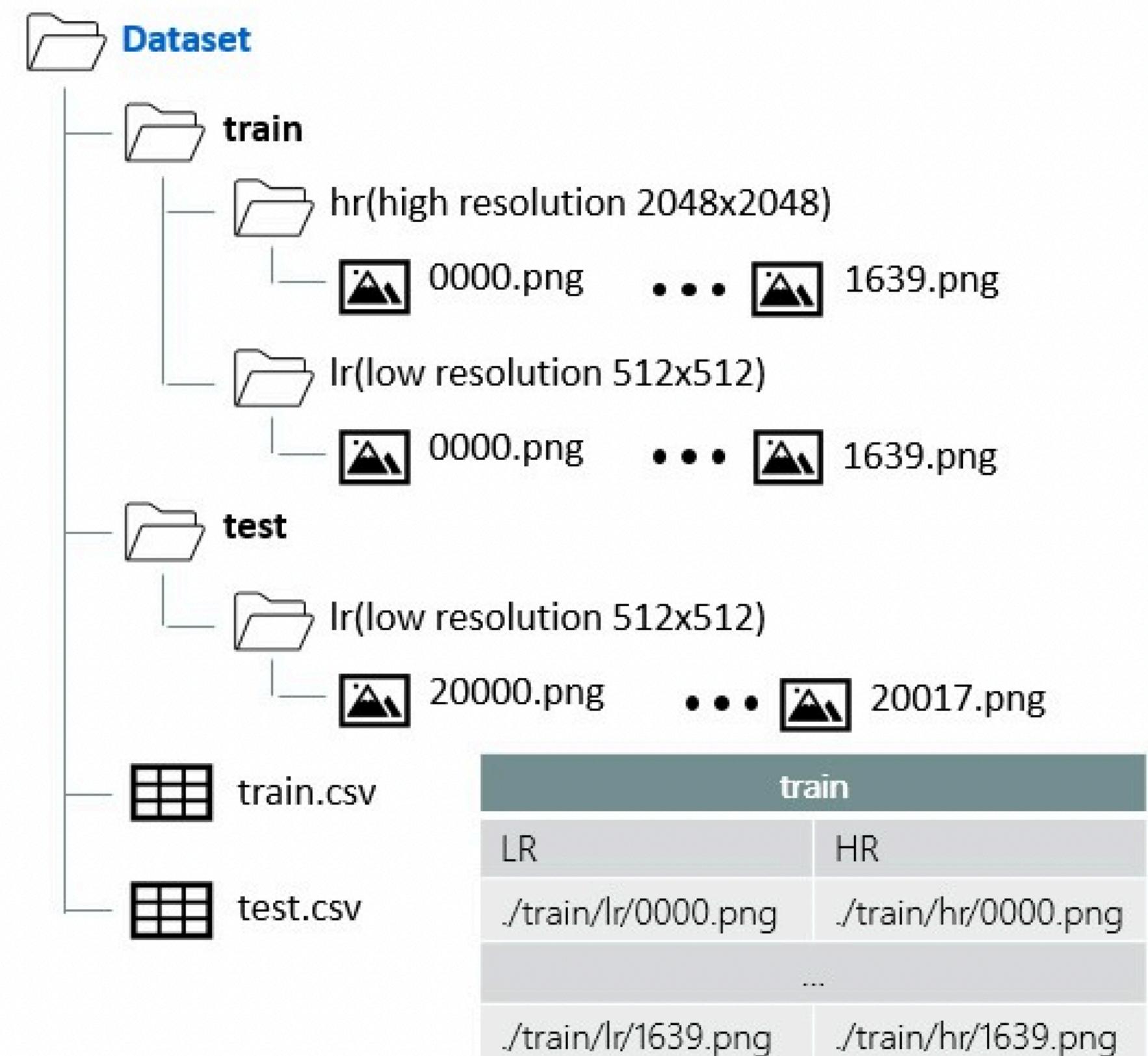


데이터 특징

데이터 특징

■ 데이터 구조

- > 고해상도(2048x2048)이미지와 저해상도(512x512) 이미지 총 1640쌍의 학습용 데이터
- > 저해상도 이미지 18개의 테스트 데이터
- > 테스트 데이터를 대상으로 초해상화를 통해 생성한 고해상도 이미지를 대회 평가용으로 제출

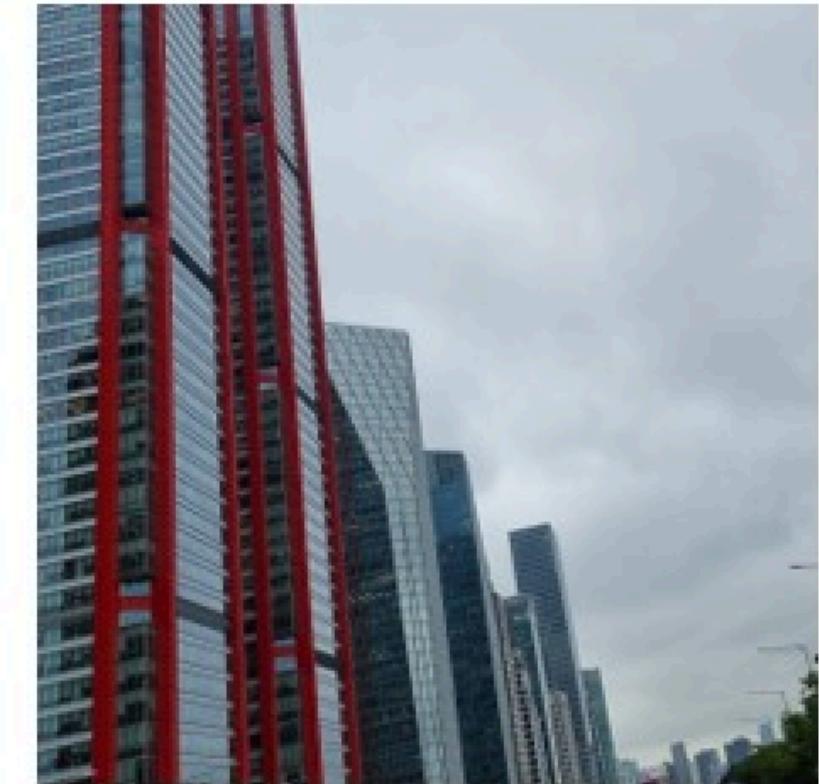
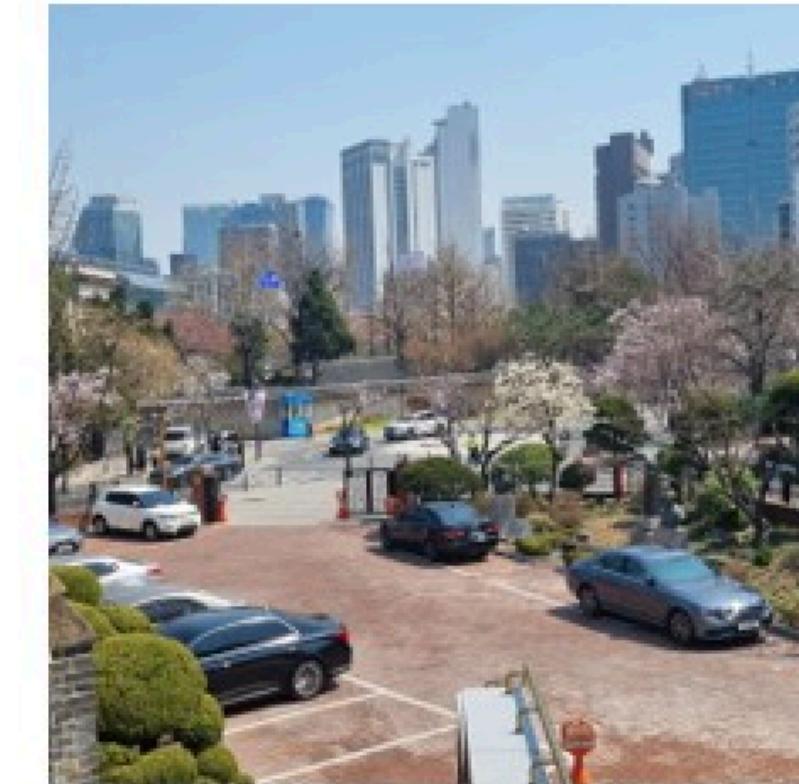


train	
LR	HR
/train/lr/0000.png	/train/hr/0000.png
...	
/train/lr/1639.png	/train/hr/1639.png

데이터 특징

■ 제공된 데이터 특징

- > 주로 수도권 도시 풍경을 촬영한 이미지
- > 저해상도 이미지는 원본 이미지를 다양한 방법으로 512x512 해상도로 감소시킨 것으로 보임
- > 빌딩, 창문 등 미세한 정보가 많아 이를 정확하게 복원하는 것이 PSNR 성능의 주요 요소로 판단



< 제공된 데이터 이미지 >

■ 유사 데이터셋

- > Urban 100(Single Image Super-Resolution)
- > Cityscape(Semantic Segmentation)



< Urban100(좌), Cityscape(우) >

MODEL SELECTION

초기 모델 선정

- 초기 모델 선정은 Papers with code의 Image Super-Resolution 관련 성능 Benchmarks를 참고 (특히, 유사 데이터 셋으로 판단한 Urban100-4x upscaling)
- 초기 선정 모델은 **HAT, SwinIR** 두 가지 모델로 선정

Rank	Model	PSNR ↑	SSIM	LPIPS	Perceptual Index	Paper	Code	Result	Year	Tags ⚡
1	<u>HAT-L</u>	28.60	0.8498			Activating More Pixels in Image Super-Resolution Transformer	🔗	📄	2022	
2	<u>HAT</u>	28.37	0.8447			Activating More Pixels in Image Super-Resolution Transformer	🔗	📄	2022	
3	<u>SwinFIR</u>	28.12	0.8393			SwinFIR: Revisiting the SwinIR with Fast Fourier Convolution and Improved Training for Image Super-Resolution	🔗	📄	2022	
4	<u>SwinIR</u>	27.45	0.8254			SwinIR: Image Restoration Using Swin Transformer	🔗	📄	2021	
5	<u>HBPN</u>	27.3	0.818			Hierarchical Back Projection Network for Image Super-Resolution	🔗	📄	2019	
6	<u>LTE</u>	27.24				Local Texture Estimator for Implicit Representation Function	🔗	📄	2021	
7	<u>SAN</u>	27.23	0.8169			Second-Order Attention Network for Single Image Super-Resolution	🔗	📄	2019	
8	<u>CSNLN</u>	27.22	0.8168			Image Super-Resolution with Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining	🔗	📄	2020	

최종 모델 선정

- 초기 검토 모델들은 Transformer 계열의 모델
- Transformer 계열의 모델은 다양하고 방대한 데이터에 대해서는 효과적이지만 데이터가 **한정적인** 경우에는 좋지 않을 것으로 판단
Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
- 따라서, CNN 계열의 RRDBNet을 Baseline으로 사용
BasicSR repository : https://github.com/XPixelGroup/BasicSR/blob/master/basicsr/archs/rrdbnet_arch.py

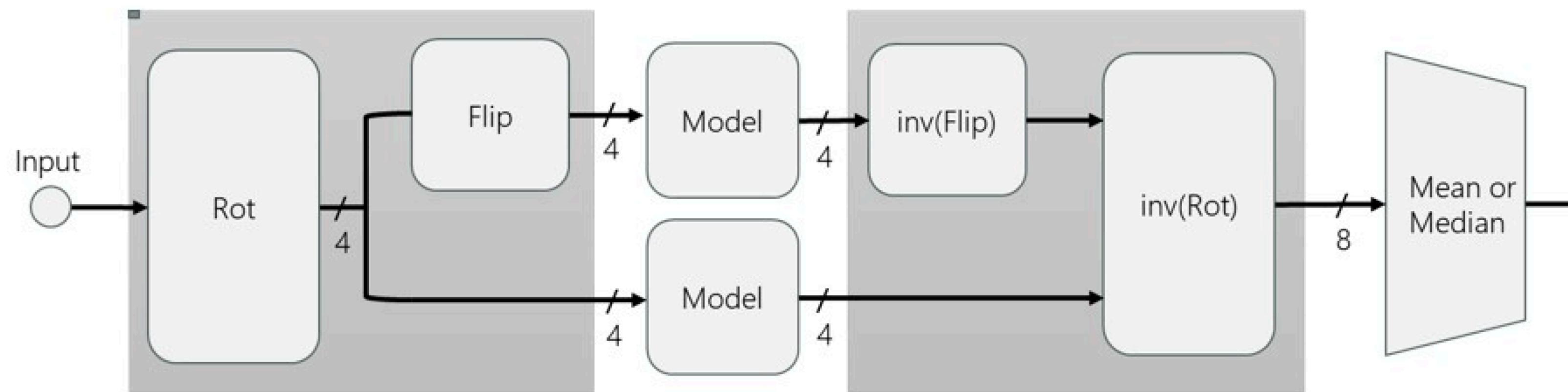
METHODS

학습방법

- **Dataset : 대회에서 주어진 데이터만 사용**
> 저해상도를 만드는 방법이 공유되지 않았기 때문에 추가 데이터는 불필요할 것으로 판단
- **Augmentation : Rotation, Flip**
> 특정 이미지 패턴에 모델이 편향되지 않도록 train 데이터에 rotation, flip 적용
- **Loss function : MSELoss**
> 성능 지표로 PSNR만 고려하기 때문에, MSELoss를 Loss function
- **MultiStepLR 적용**
> 200,000 iter마다 learning rate가 0.5배가 되도록 설정
- **Post-Processing : Geometric Self-Ensemble**
> 후처리 과정으로 Geometric Self-Ensemble을 이용하여 성능 향상

GEOMETRIC SELF-ENSEMBLE

- EDSR Paper에서 언급된 ‘NTIRE 2017 SISR challenge’를 우승할 때 사용했던 방법인 Geometric Self-Ensemble 방법 채용. [1]

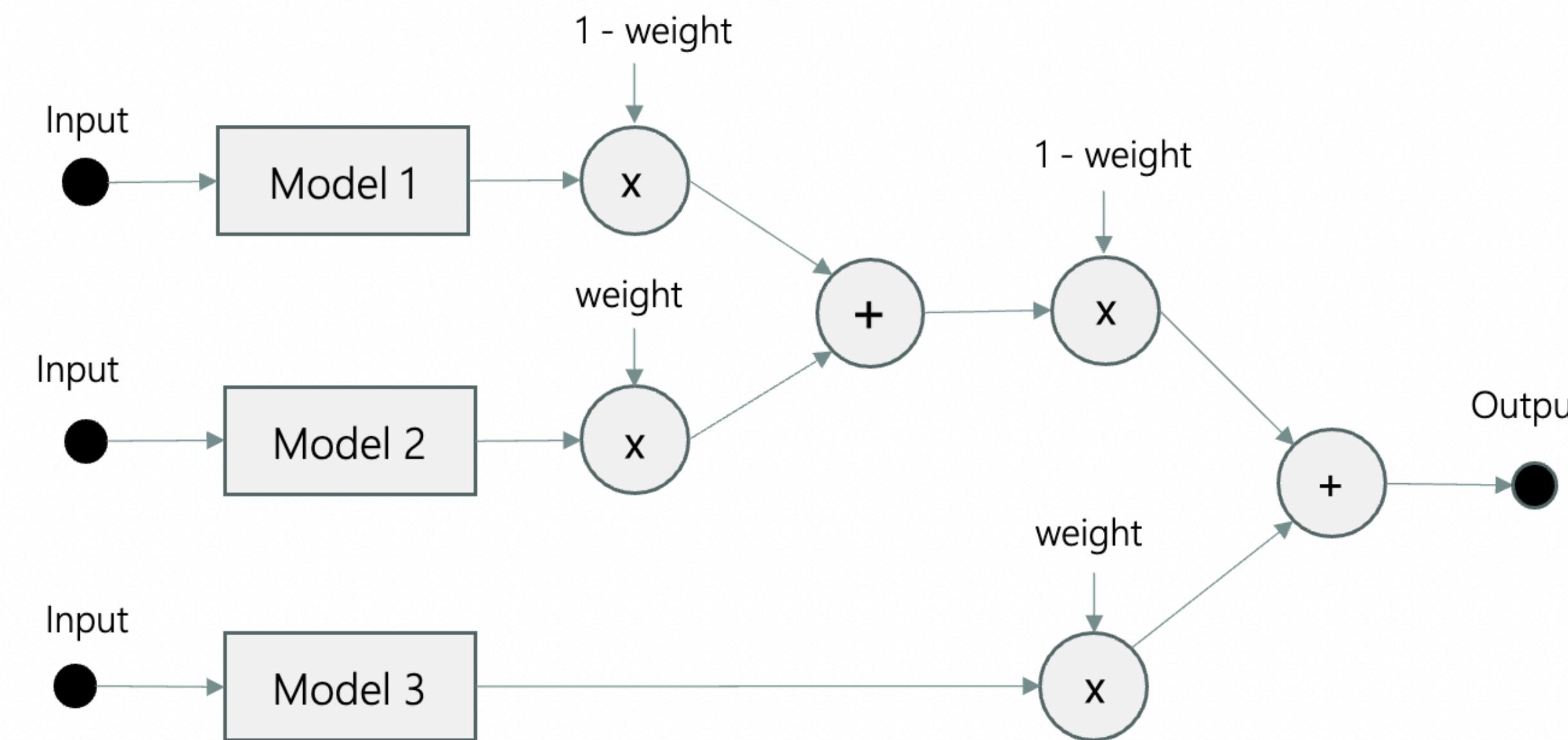


- Rotation, Flip을 이용하여 학습된 모델에 좌우 반전, 4가지 회전(0도, 90도, 180도, 270도)을 적용한 이미지 8개를 통과시켜 나온 결과들을 평균 혹은 중앙값을 취하여 성능을 개선시킴.
(최종 모델은 평균 값 이용)

[1] Lim, Bee, et al. "Enhanced deep residual networks for single image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.

(ADDITIONAL) SOFT VOTING

- 팀원들의 각각의 모델의 validation PSNR값에 근거하여 weight를 설정하여 soft voting 수행



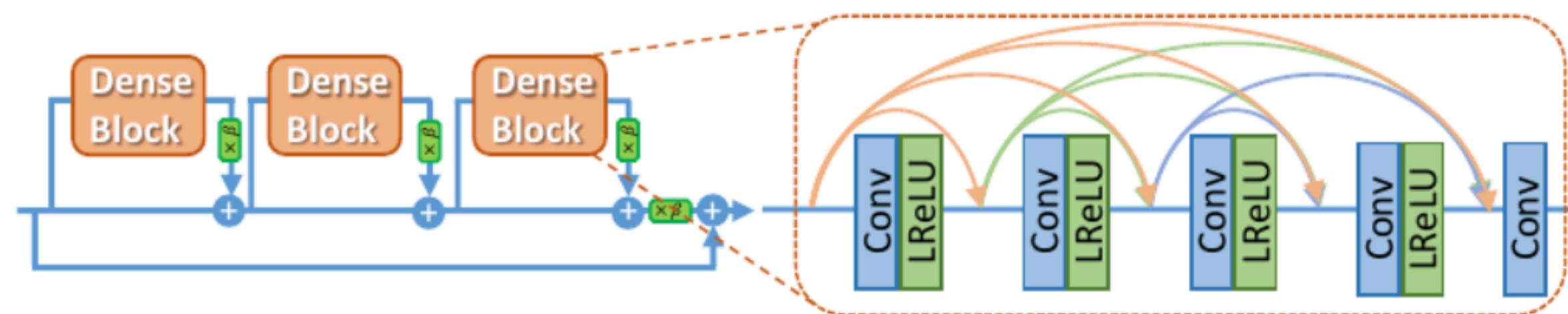
> Voting은 각 모델 별 성능 편차가 존재하여 큰 개선 효과가 없었음

NETWORK ARCHITECTURE

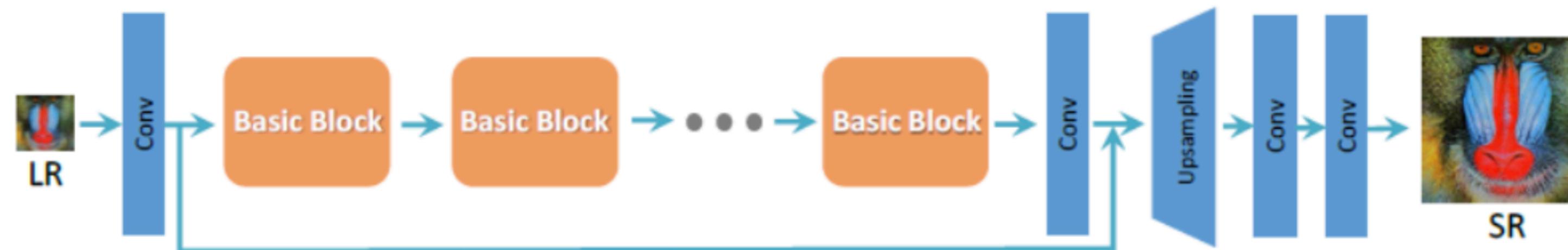
■ Baseline으로 RRDBNet[1]을 이용

> RRDB : 5개의 Conv.로 이루어진 dense block을 3번 통과 후 통과 전 이미지와 weighted sum

Residual in Residual Dense Block (RRDB)



> RRDBNet Architecture (Basic Block : RRDB)



[1] Wang, Xintao, et al. "Real-esrgan: Training real-world blind super-resolution with pure synthetic data." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

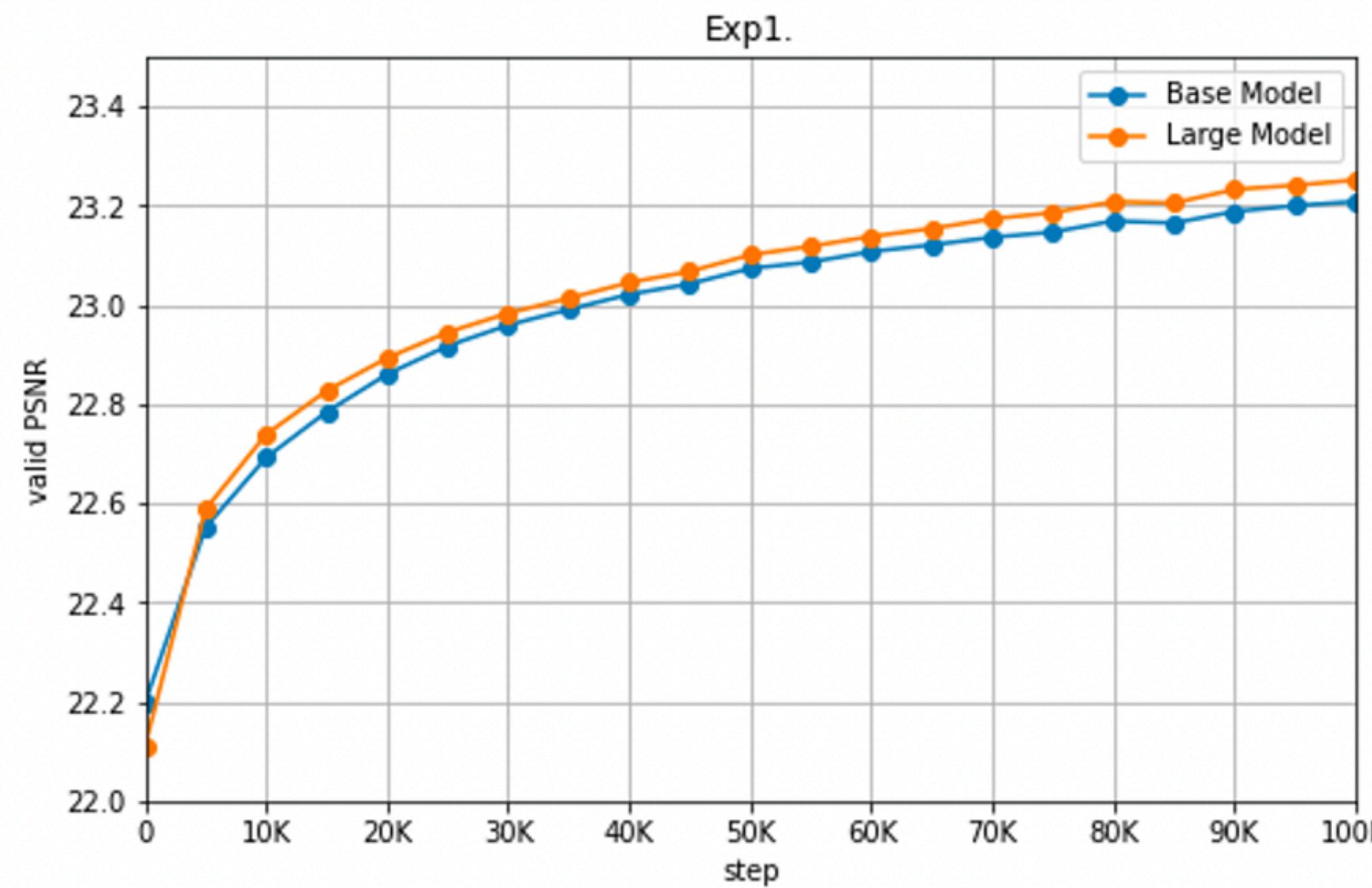
EXPERIMENTS

EXPERIMENTS

- RRDBNet 기본 architecture를 baseline으로 아래와 같은 실험들을 진행
 1. Model hyper-parameters 조정 (num_feat, num_grow_ch)
 2. 이미지 Rotation 모델
 3. Body(RRDB)의 위치 조정
 4. Gated Convolution 사용
 5. Pretrained Model 사용

1. MODEL HYPER-PARAMETERS 조정 (LARGE MODEL)

- num_feat : network 전체에서 통용되는 feature map의 개수
- num_grow_ch : dense block 내부에서 증가되는 channels의 개수

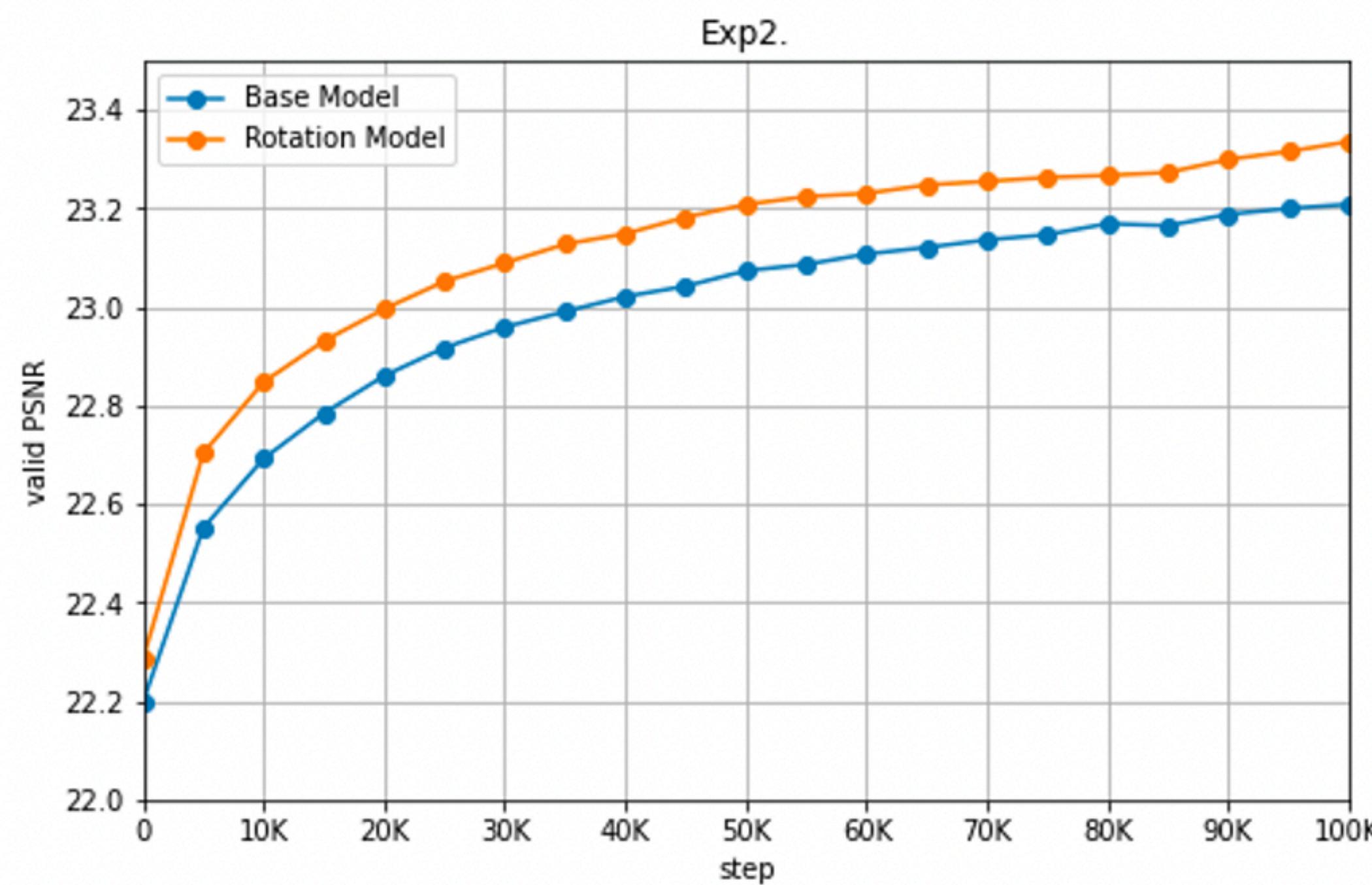


	num_feat	num_grow_ch	100k steps PSNR
Base Model	64	32	23.199
Large Model	96	48	23.241

> Large model이 base model에 비해 근소하게 PSNR 성능이 좋으나, training 속도가 1.4배 정도 느리기 때문에 base model이 빠르게 좋은 결과를 볼 수 있을 것이라 판단

2. 이미지 ROTATION 모델

- 0도, 90도, 180도, 270도 이미지를 회전한 후 network에 독립적으로 입력하여 낳은 결과를 평균 내어 최종 결과로 출력하는 네트워크 구성

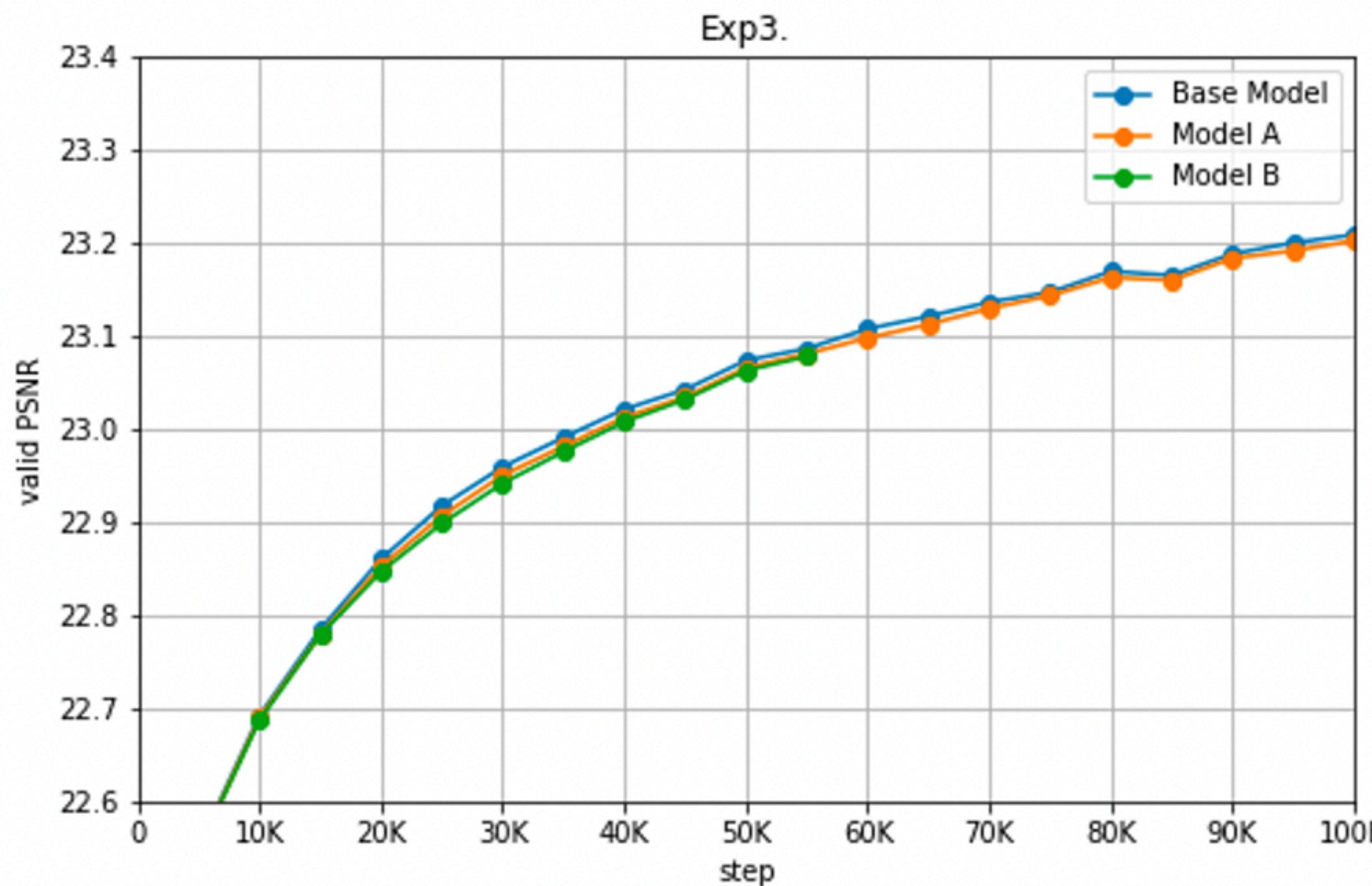


100k steps PSNR	
Base Model	23.199
Large Model	23.315

> Rotation model이 base model에 비해서 PSNR 수치가 상당히 높으나 트레이닝 속도가 약 3.5배 차이나 대회 시간 내에 좋은 결과를 얻기 힘들 것이라 판단

2. BODY(RRDB)의 위치 조정

- RRDBNet은 I-R-U-U-O의 구조를 기본적으로 가짐 (Input-RRDB-Upsample-Upsample-Output)
- Upsample 이후 RRDB를 추가하는 구조를 실험



> Base model : 기본적인 RRDBNet I-R-U-U-O

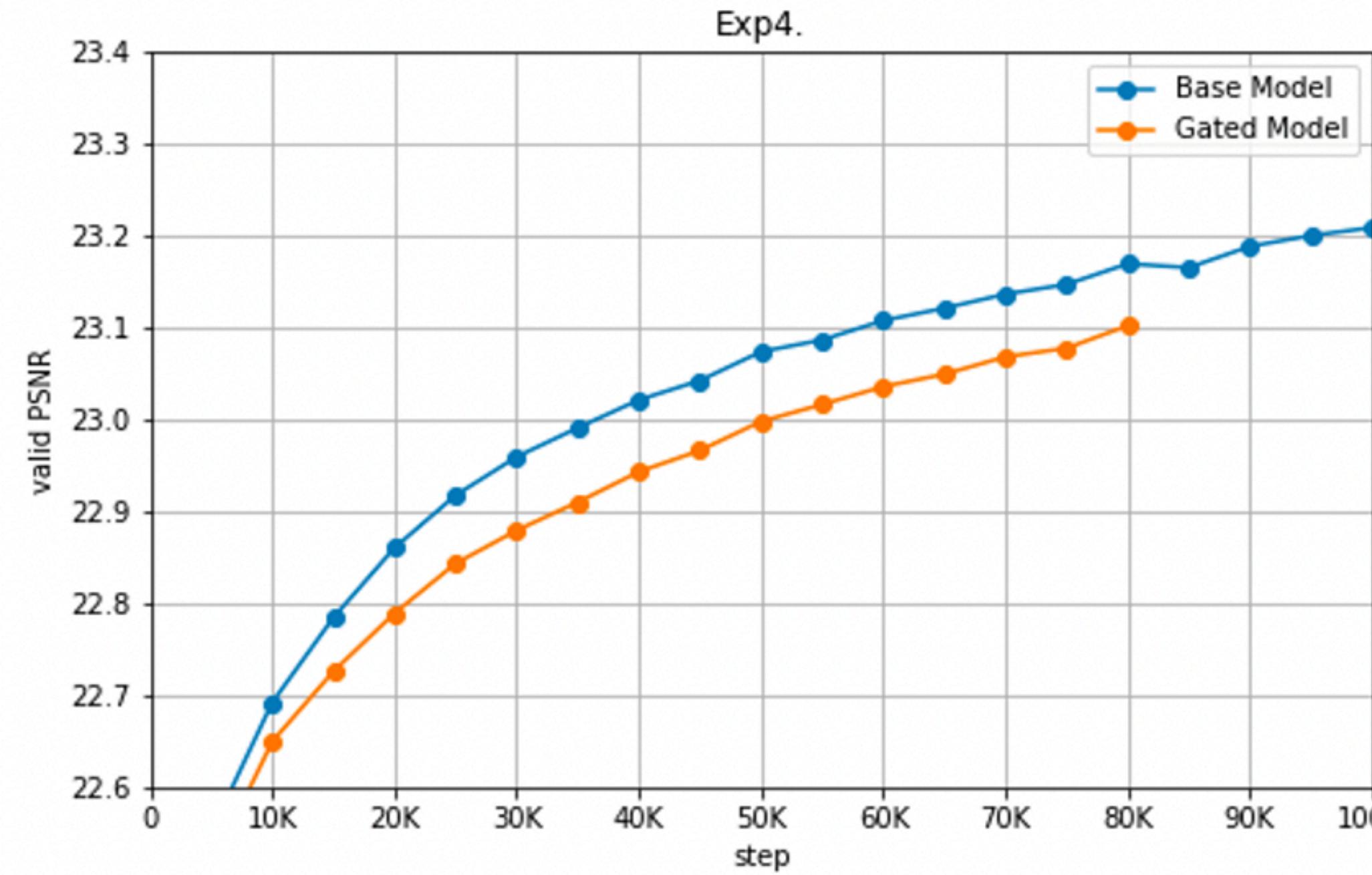
Model A : I-R-U-R-U-O 구조

Model B : I-R-U-R-U-R-O 구조

> Valid PSNR의 결과가 base model 대비 차이가 없거나, 오히려 조금 떨어짐

4. GATED CONVOLUTION 사용

- 기본 Convolution module을 Wavenet[1], PixelCNN++[2]에 사용된 gated convolution으로 변경



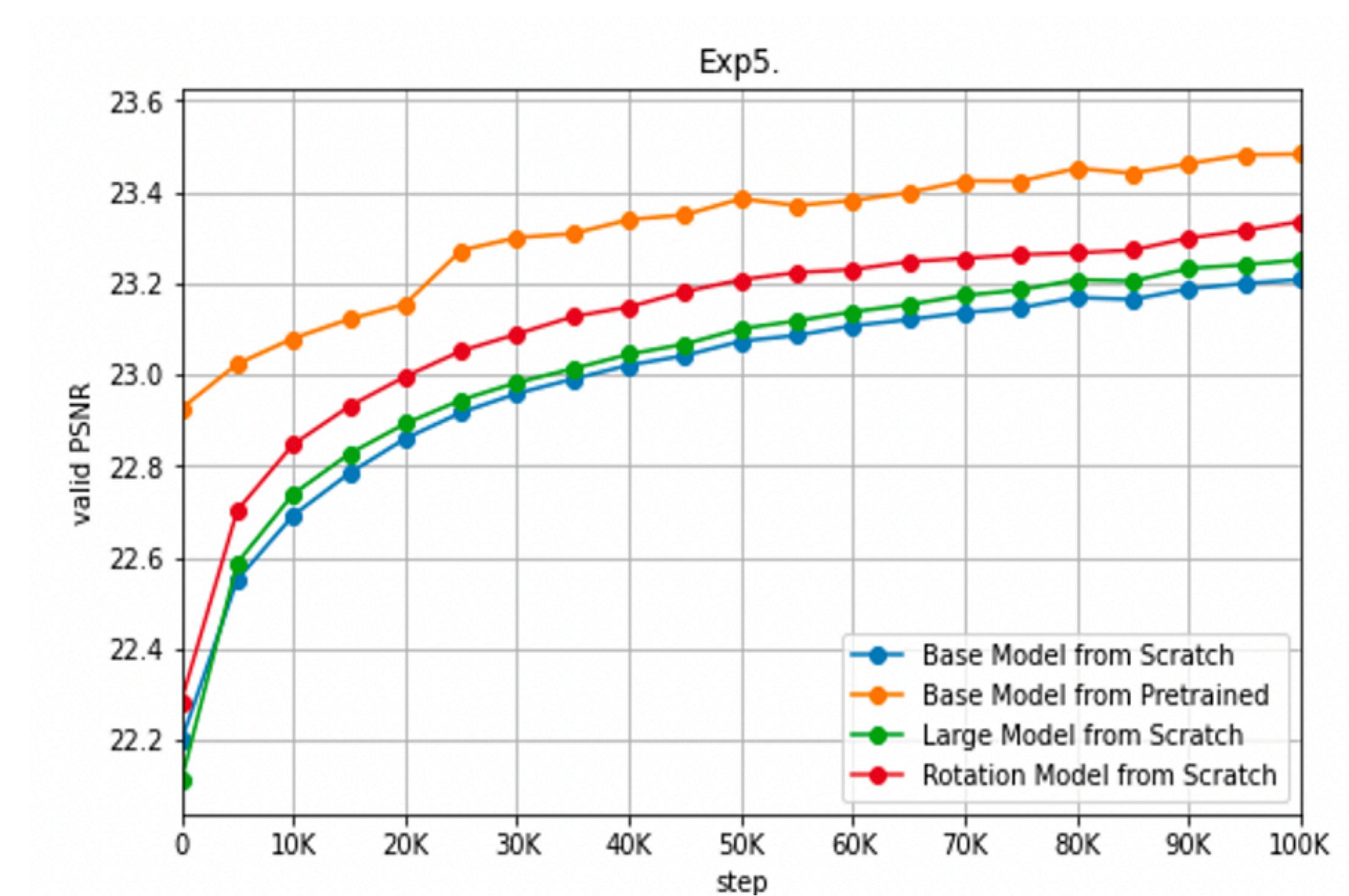
> Gated Convolution을 이용한 모델이 base model
에 비해서 확연하게 성능이 떨어지는 것을 확인

[1] Oord, Aaron van den, et al. “Wavenet: A generative model for raw audio.” arXiv preprint arXiv:1609.03499 (2016)

[2] Salimans, Tim, et al. “Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications.” arXiv preprint arXiv:1701.05517 (2017)

5. PERTAINED MODEL 사용

- Real-ESRGAN[1]에서 애니메이션, 영화 등의 데이터로 학습한 pertained model을 시작으로, 대회에서 제공한 dataset으로 fine-tuning 진행

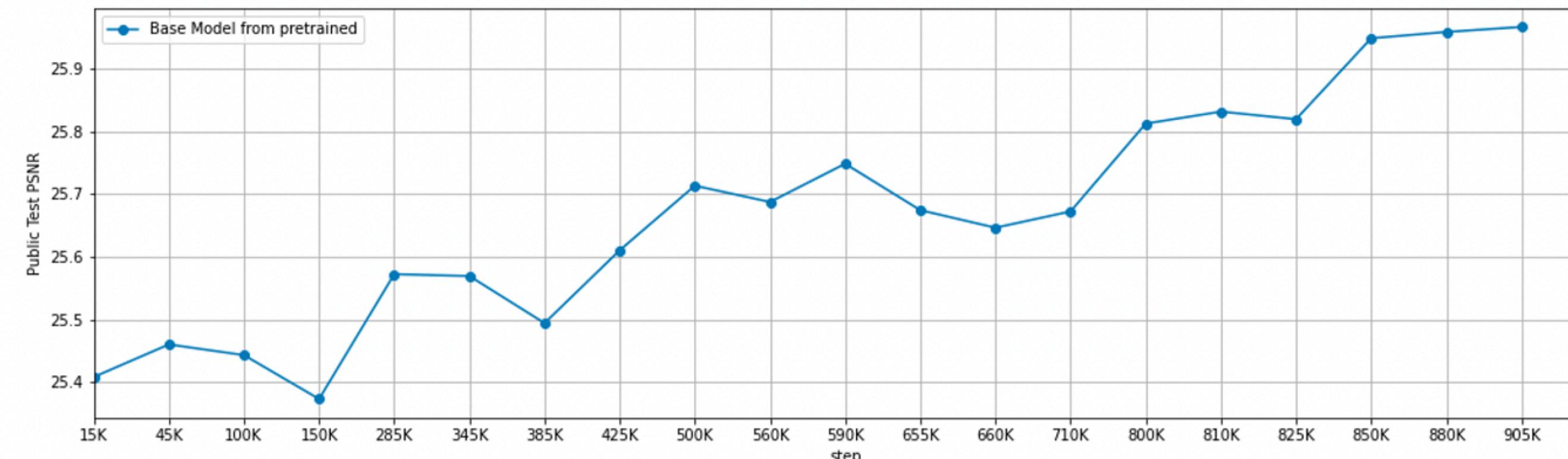


	100k steps PSNR
Base Model from Scratch	23.199
Base Model from Pretrained	23.481
Large Model from Scratch	23.241
Rotation Model from Scratch	23.315

> Pretrained model을 이용한 학습이 앞서 보았던 여러 방법들 (large model, rotation model)들 보다도 빠르게 좋은 성능을 보이는 것을 확인

CONCLUSION

- Model의 hyper-parameters를 조정한 large model과 이미지를 rotation하여 network에 넣고 결과에 mean을 위한 ensemble model이 base model 대비 좋은 성능을 가지는 것을 확인
- 그러나 위 모델들은 base model 대비 트레이닝 시간이 길고, 마침 pretrained model이 있어 이것을 시작으로 fine-tuning 하여 더 빠르게 트레이닝을 진행할 수 있었다.
- training을 시작하고 905K번째의 checkpoint에서 가장 높은 Public test PSNR을 가져왔으며, 이를 최종 제출로 정하였다.



감사합니다.