

# Stochastic Weight Averaging 알고리즘을 이용한 이미지 초해상도 성능 개선

윤정환, 조남익

서울대학교 전기정보공학부

hamil951753@snu.ac.kr, nicho@snu.ac.kr

## Stochastic Weight Averaging for Improving the Performance of Image Super-Resolution

Jeong Hwan Yoon, Nam Ik Cho

Department of ECE, Seoul National University

### 요 약

단일 이미지 초해상도는 딥러닝의 발전과 함께 놀라운 성능 향상이 이루어 졌다. 이러한 딥러닝 모델은 매우 많은 파라미터를 갖고 있어 많은 연산량과 메모리를 필요로 한다. 하지만 사용할 수 있는 리소스는 한정되어 있기 때문에 네트워크를 경량화 시키려는 연구도 지속되어 왔다. 본 논문에서는 Stochastic Weight Averaging (SWA) 알고리즘을 이용하여 상대적으로 적은 양의 메모리와 연산을 추가해 이미지 초해상도 모델의 성능을 높이고 안정적인 학습을 달성하였다. SWA 알고리즘을 적용한 모델은 그렇지 않은 모델에 비해 테스트셋에서 최대 0.13dB의 성능 향상을 보였다.

### 1. 서론

고해상도 디스플레이의 발전으로 이미지나 영상의 화질을 개선하는 기술에 대한 수요가 늘어나고 있다. 그 중에서도 단일 이미지 초해상도 (Single Image Super-Resolution)는 주어진 저해상도 이미지를 이용해 고해상도 이미지를 추정하는 과제이다. 저해상도 이미지의 한 픽셀은 고해상도 이미지에서의 픽셀 여러 개와 대응되기 때문에 이미지 초해상도는 일 대 다 대응 문제이며 정답이 유일하지 않은 ill-posed inverse 문제로 분류된다.

최근 합성곱 신경망의 등장으로 이미지 초해상도 분야에서도 합성곱 신경망을 이용한 학습 기반의 모델 들을 중심으로 눈에 띄는 성능 향상이 일어났다. SRCNN [1]은 이

미지 초해상도 모델에 처음으로 딥러닝 방법을 적용하여 기존의 전통적인 방식을 이용한 이미지 초해상도 연구들보다 훨씬 뛰어난 성능을 보였다. EDSR [2]은 잔차 신경망 구조 (residual network)를 이용하여 네트워크를 깊게 쌓으면서도 학습이 가능하게 하였다. RCAN [3]은 채널 어텐션 매커니즘을 이용하여 추가적으로 성능을 개선시켰다. SAN [4]은 채널 어텐션 매커니즘 뿐만 아니라 이미지 피쳐 사이의 전역적인 관계까지 고려하여 우수한 성능을 달성하였다.

이러한 기존의 연구들은 합성곱 신경망 모델의 구조를 변화시켜가며 이미지 초해상도의 성능을 향상시켰다. 하지만 단순히 모델의 깊이를 증가시키거나 구조를 변화시키는 방법들에는 한계가 존재했다. 게다가 깊은 모델들은 많은 수의 파라미터를 갖고 있기 때문에 학습과 추론 과정에서 많은 양의 연산과 리소스를 필요로 한다. 사용할 수 있는 리소스는 한정되어 있기 때문에 모델을 경량화 시키려는

연구도 계속되어 왔다 [5].

본 논문에서는 학습과정에서 Stochastic Weight Averaging (SWA) [6] 알고리즘을 추가하여 학습과정을 더욱 안정적으로 만들고 이미지 초해상도 모델의 성능을 개선하고자 하였다. SWA 알고리즘을 이용한 모델은 추가적인 연산과 메모리 사용량이 크지 않음에도 테스트셋에서 최대 0.13dB의 PSNR 상승을 달성할 수 있었다.

## 2. 실험

### 2.1 Stochastic Weight Averaging 알고리즘

기존의 확률적 경사 하강법 (Stochastic Weight Descent) 기반의 학습 방식들은 일정 주기마다 가중치들을 업데이트 한다. SWA 알고리즘은 이에 더해 일정 주기마다 모델의 가중치들의 가중합 (weighted sum)을 계산한다. 식 (1)은 모델의 가중치들을 업데이트하는 알고리즘이다. 이때  $w_{swa}$  와  $w$ ,  $n_{models}$  는 각각 SWA 를 거친 가중치들의 가중합, optimizer 를 통해 업데이트된 가중치, SWA 모델의 수를 의미한다 [6].

$$\frac{w_{swa} \cdot n_{models} + w}{n_{models} + 1} \rightarrow W_{swa} \quad (1)$$

이러한 SWA 알고리즘을 추가하면 하나의 모델을 이용해 시간축으로 앙상블 하는 효과를 낼 수 있다. 따라서 상대적으로 적은 양의 연산과 메모리를 사용해 모델의 일반화 성능을 높일 수 있고, 학습과정에서도 더욱 안정적인 수렴이 가능하게 한다.

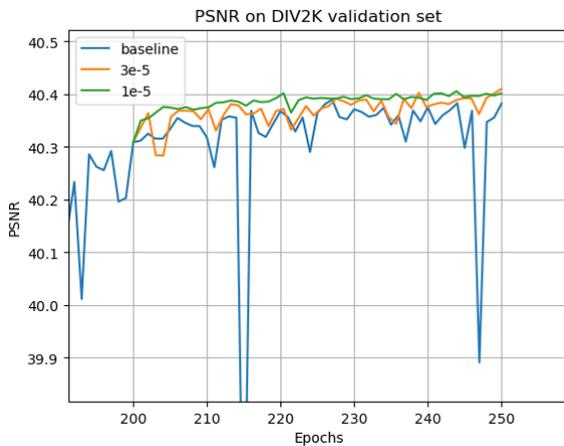


그림 1. DIV2K validation dataset 에 대한 PSNR 변화.

### 2.2 실험 내용

본 논문에서는 SWA 알고리즘 기반의 모델 앙상블을 이미지 초해상도 분야에서 적용하기 위해 RCAN [3]의 구조를 기반으로 한 실험을 진행하였다. 구체적으로, 기존의 방법대로 모델을 사전학습 시킨 뒤 SWA 알고리즘을 적용해 추가적인 학습을 진행하였다. 본 논문에서는 SWA 가중합을 구하는 주기를 1 epoch 으로 설정하여 매 epoch 마다 optimizer 를 통해 업데이트 된 가중치들의 가중합을 계산해 주었다.

표 1 다양한 학습률과 SWA 알고리즘 사용에 대한 초해상도 복원 결과. SWA 모델은 200 번의 반복 이후 각각 0.00005, 0.00003, 0.00001 의 학습률을 사용함. 각 테스트 셋에 대하여 가장 좋은 결과는 굵은 글씨로, 두번째로 좋은 결과는 밑줄로 표시함.

	Set5	Set14	B100	Urban100
Baseline	<u>38.198</u>	33.891	32.327	32.849
SWA_1	38.180	33.922	32.320	32.875
SWA_2	<b>38.200</b>	<u>33.943</u>	<u>32.334</u>	<u>32.940</u>
SWA_3	38.188	<b>33.948</b>	<b>32.339</b>	<b>32.977</b>

### 2.3 실험 결과

실험은 DIV2K 데이터셋 800장을 이용하여 학습을 진행하였고, Set5, Set14, B100, Urban100 에 대해 PSNR 을 측정해 평가를 진행하였다. 학습은 ADAM optimizer ( $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ )를 활용하였다.

Baseline 모델의 경우 200 epoch 까지는 0.0001 의 학습률로 학습을 진행하였고 마지막 50 epoch 은 학습률을 절반으로 낮춰 모델을 학습시켰다. SWA 모델의 경우 200 epoch까지 학습시킨 baseline 모델을 SWA 알고리즘을 적용하여 추가적으로 50 epoch 씩 학습시켰다. 표 1 의 결과를 보면, SWA 알고리즘을 적용한 모델이 baseline 모델에 비해 최대 0.13dB 의 성능 향상을 내는 것을 확인할 수 있다. SWA 알고리즘에서 너무 큰 학습률은 오히려 학습에 방해가 되기 때문에 적절한 값을 설계해 주어야 함을 알 수 있다.

또한, 그림 1 에서 알 수 있듯이, SWA 알고리즘을 적용

한 모델이 그렇지 않은 모델에 비해서 더욱 안정적으로 수렴하며 더 좋은 성능을 낸다. 이는 학습과정에서 모델의 가중치들을 평균을 내었기 때문에 더욱 일반적인 정답을 탐색하고 있다고 추론할 수 있다.

[6] Izmailov, Pavel, et al. "Averaging weights leads to wider optima and better generalization." *arXiv preprint arXiv:1803.05407* (2018).

### 3. 결론

본 논문에서는 Stochastic Weight Averaging 알고리즘을 활용하여 기존의 모델에 매우 적은 양의 연산을 추가하여 모델의 일반화 성능을 개선하는 연구를 진행하였고, 일부 테스트셋에 대해 성능이 향상된 것을 확인할 수 있었다.

### 감사의 글

이 논문은 삼성전자의 지원과 2021 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2021R1A2C2007220).

### 참고문헌

[1] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015): 295-307.

[2] 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.

[3] Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

[4] Dai, Tao, et al. "Second-order attention network for single image super-resolution." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[5] Wang, Xuehui, et al. "Lightweight single-image super-resolution network with attentive auxiliary feature learning." *Proceedings of the Asian Conference on Computer Vision*. 2020.