초해상화 모델의 활성함수 변경에 따른 성능 분석

유영준, 김대희, 이재구*
국민대학교 컴퓨터공학과
*jaekoo@kookmin.ac.kr

Performance Analysis of Various Activation Functions in Super Resolution Model

YoungJun Yoo, DaeHee Kim, JaeKoo Lee*
Dept. of Computer Science, Kookmin University

요 약

ReLU(Rectified Linear Unit) 함수는 제안된 이후로 대부분의 깊은 인공신경망 모델들에서 표준 활성함수로써 지배적으로 사용되었다. 이후에 ReLU를 대체하기 위해 Leaky ReLU, Swish, Mish 활성함수가 제시되었는데, 이들은 영상 분류 과업에서 기존 ReLU 함수 보다 향상된 성능을 보였다. 따라서 초해상화(Super Resolution) 과업에서도 ReLU를 다른 활성함수들로 대체하여 성능 향상을 얻을 수 있는지 실험해볼 필요성을 느꼈다. 본 연구에서는 초해상화 과업에서 안정적인 성능을 보이는 EDSR(Enhanced Deep Super-Resolution Network) 모델의 활성함수들을 변경하면서 성능을 비교하였다. 결과적으로 EDSR의 활성함수를 변경하면서 진행한 실험에서 해상도를 2배로 변환하는 경우, 기존 활성함수인 ReLU가 실험에 사용된 다른 활성함수들 보다 비슷하거나 높은 성능을 보였다. 하지만 해상도를 4배로 변환하는 경우에서는 Leaky ReLU와 Swish 함수가 기존 ReLU 함수 대비 다소 향상된 성능을 보임을 확인하였다. 구체적으로 Leaky ReLU를 사용했을 때 기존 ReLU보다 영상의 품질을 정량적으로 평가할 수 있는 PSNR과 SSIM 평가지표가 평균 0.06%, 0.05%, Swish를 사용했을 때는 평균 0.06%, 0.03%의 성능 향상을 확인할 수 있었다. 4배의 해상도를 높이는 초해상화의 경우, Leaky ReLU와 Swish가 ReLU 대비 향상된 성능을 보였기 때문에 향후 연구에서는 다른 초해상화 모델에서도 성능 향상을 위해 활성함수를 Leaky ReLU나 Swish로 대체하는 비교실험을 수행하는 것도 필요하다고 판단된다.

1. 서론

초해상화(Super Resolution)는 저해상도의 영상을 고해상도 영상으로 변환하는 과업이다. 기존 영상의 품질을 향상할 수 있는 특성 때문에 초상해도는 CCTV 영상 데이터에서의 얼굴 인식, 특정 장면에서의 객체 탐지, 천문 영상, 의학 촬영 영상 등의 다양한 분야에서 사용된다.[1] 또한 초해상화는 크게 하나의 영상을 이용하는지, 여러 영상을 이용하는지에 따라 SISR(Single Image Super Resolution)과 MISR(Multi Image Super Resolution)로 나뉜다. MISR이 SISR에 기초를 두고 있기 때문에 주로 SISR에 대한 연구가 주를 이루고 있다.

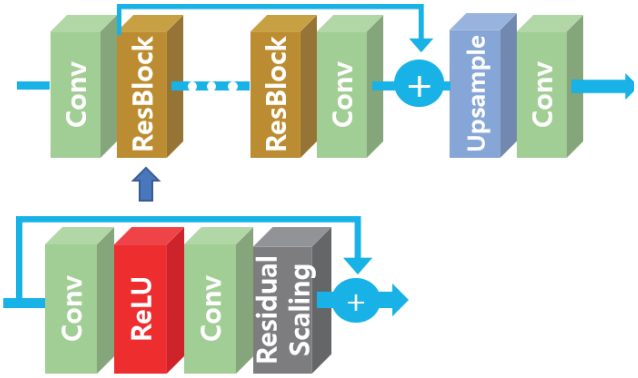
EDSR(Enhanced Deep Super-Resolution Network) [2]은 NTIRE2017 Super-Resolution Challenge[3]에서 우승한 초해상화 모델인데 제안된 당시 SOTA(State-Of-The-Art)를 달성하였다. EDSR 이전에 제시되었던 SRResNet[4]은 영상 분류 과업과 같은 고수준의 과업을 해결하기 위해 고안된 ResNet[5]이라는 인공신경망 구조를 사용하였다. 하지만 SISR과 같은 저수준 과업을 풀기 위해서 ResNet 구조를 동일하게 사용하는 것은 문제가 있었다. EDSR는 SRResNet 보다 성능이 향상되었는데, SRResNet과 같은 이전 인공신경망 구조들에서 불필요한 모듈들을 제거하여 최적화를 했고 훈련 과정을 안정화였다. 최적화를 통해 확보한 메모리 공간을 사용해 모델의 크기를 늘려서 구조를 보다 적합하게 수정할 수 있었다. 활성함수로는 ReLU[6]를 사용하였다.

최근 깊은 인공신경망 연구에서 ReLU 함수는 다른 활성함수들과 비교해서 간단한 구현과 일관된 성능 때문에 대부분의 모델들에서 표준 활성함수로써 지배적으로 사용되었다. 특히 영상 분류 과업에서 좋은 성능을 보였다. 최근 ReLU의 단점들을 보완한 다양한 활성함수들이 제안되었으며 대표적으로 Leaky ReLU[7], Swish[8], Mish[9]가 있다. 이 활성함수들은 특히 영상 분류 과업에서 기존의 ReLU 함수보다 좋은 성능을 보였다. 따라서 초해상화 과업에서도 활성함수를 대체하여 이러한 성능 향상을 얻을 수 있는지 검증할 필요성이 대두되었다.

본 논문에서는 초해상화 모델의 활성함수를 기존의 ReLU 에서 Leaky ReLU, Swish, Mish 로 대체했을 때 모델의 성능을 높일 수 있는지 확인하였다. 초해상화 모델로는 초해상화 평가지표인 PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity)에서 안정적인 성능을 보이는 EDSR 을 사용하였다. 기존 ReLU 와 더불어 Leaky ReLU, Swish, Mish 활성함수를 EDSR 에 변경하면서 활성함수에 따른 성능을 비교 및 분석하였다.

실험에서는 초해상화 과업에서 주로 사용되는 5 개의 데이터집합을 사용하였고, 해상도를 2 배로 변환하는 경우와 해상도를 4 배로 변환하는 경우로 나누어, 각 모델에서 활성함수에 따른 성능 비교를 진행하였다.

2. EDSR (Enhanced Deep Super-Resolution Network)

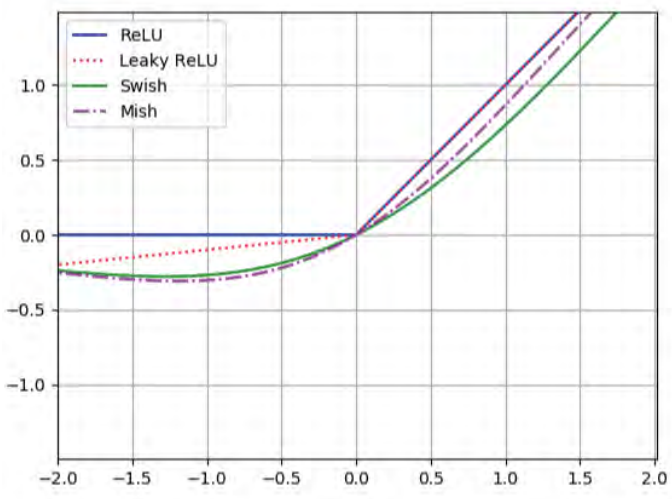


(그림 1) EDSR 구조

EDSR 은 SRResNet 이 잔차 블록(Residual Block)에서 사용했던 배치 정규화(Batch Normalization)층을 모두 제거하였다. 인공신경망의 특징들을 정규화시켜 인공신경망의 유연성을 저해하는 배치 정규화 층 제거를 통해 성능을 향상할 수 있었다. 또한 배치 정규화 층을 제거함으로써 GPU 메모리 사용량을 줄였는데, 이렇게 확보한 메모리를 사용해 더 큰 모델을 구성하여 제한된 계산 자원하에서 SRResNet 보다 더 좋은 성능을 얻을 수 있었다. 활성함수로는 ReLU 함수를 사용하였는데, (그림 1)에서 보면 잔차 블록에서 ReLU 를 사용하는 것을 볼 수 있다.

3. 활성함수 (Activation Function)

(그림 2)에서는 본 논문에서 사용된 ReLU, Leaky ReLU, Swish, Mish 활성함수를 표현하였다. 식 (1)은 ReLU 함수의 수식을 표현하였으며, ReLU 를 사용하면 깊은 인공신경망에서 Sigmoid 또는 Tanh 함수를 사용한 것보다 더 쉽게 최적화가 된다. Sigmoid 와 Tanh 같은 활성함수들은 입력이 커지면 미분 값이 포화가 되어 오류를 전달 못 하는 문제가 있었는데, ReLU 는 양수 영역에서 값이 선형적으로 출력되어 경사소멸(Vanishing Gradient)을 피할 수 있다. 그리고 구현이 매우 간단하고 연산 비용이 많이 들지 않는다.



(그림 2) 활성함수 비교 그래프

$$f(x) = \begin{cases} (x < 0) & f(x) = 0 \\ (x \geq 0) & f(x) = x \end{cases} \quad (1)$$

또한 입력으로 들어오는 값이 음수면 출력을 모두 0 으로 처리하여 규제 효과를 얻을 수 있다. 하지만 이러한 특성 때문에 해당 노드가 학습이 안 되는 단점이 있다.

$$f(x) = \begin{cases} (x < 0) & f(x) = \alpha x \\ (x \geq 0) & f(x) = x \end{cases} \quad (2)$$

이러한 단점을 극복하고자 Leaky ReLU 라는 활성함수가 제안되었다. 식 (2)는 Leaky ReLU 함수의 수식을 표현하였다. Leaky ReLU 는 음수 영역을 모두 0 으로 처리하는 ReLU 의 단점을 보완하여 만들어진 활성함수이다. Leaky ReLU 는 ReLU 와 거의 유사하지만, ReLU 와 달리 음수 영역에서, 0 에서 1 의 범위를 갖는 사용자 매개변수 α 값에 따라 조정되는 작은 기울기를 부여한다.

$$f(x) = x \cdot \sigma(x) \quad (3)$$

ReLU 의 단점을 해결하기 위해 또 다른 활성함수인 Swish 가 제안되었다. 식 (3)은 Swish 함수의 수식을 표현하였으며, Swish 는 ReLU 와 달리 부드럽고 비 단조함수(Non-monotonic Function)이다. Swish 는 큰 음수값이 입력되면 출력이 0 이 되어 미분 값이 포화되지만, 작은 음수값이 입력되면 들어온 값을 어느 정도 보존한다. 이런 특성을 바탕으로 기존 ReLU 의 문제를 어느 정도 해결하여 최적점에 더욱 잘 도달하게 해준다. 또한 Swish 의 부드러운 오차 손실 경사(Loss Landscape)는 학습률과 초깃값에 대한 민감성을 줄일 수 있다. 이후에 Swish 와 유사한 Mish 라는 활성함수도 제안되었다.

<표 1> 각 활성함수에 따른 EDSR 성능 비교 결과

Scale	DataSet	ReLU(기준)		Leaky ReLU		Swish		Mish	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
x2	DIV2K[3]	**36.56**	**0.9407**	33.71	0.9140	36.47	0.9401	36.48	0.9401
	Set5[10]	**38.19**	0.9445	35.32	0.9237	**38.19**	**0.9447**	38.19	**0.9447**
	Set14[11]	**33.95**	**0.8969**	31.60	0.8723	33.88	0.8966	33.83	0.8965
	B100[12]	**32.35**	**0.8928**	30.65	0.8683	32.31	0.8922	32.33	0.8922
	Urban100[13]	**32.97**	**0.9268**	28.02	0.8583	32.77	0.9252	32.79	0.9251
	Average	**34.80**	**0.9203**	31.86	0.8873	34.73	0.9198	34.72	0.9197
x4	DIV2K	30.73	0.8269	**30.74**	**0.8271**	**30.74**	0.8270	30.69	0.8258
	Set5	32.48	0.8686	**32.55**	**0.8692**	32.51	0.8690	32.48	0.8679
	Set14	28.82	0.7511	28.82	**0.7514**	**28.83**	0.7513	28.79	0.7501
	B100	27.72	0.7189	**27.73**	**0.7191**	**27.73**	0.7190	27.70	0.7176
	Urban100	26.65	0.7836	**26.67**	**0.7844**	**26.67**	0.7841	26.55	0.7809
	Average	29.28	0.7898	**29.30**	**0.7902**	**29.30**	0.7900	29.24	0.7885

$$f(x) = x * tanh\,(\varsigma(x)) \qquad (4)$$

식(4)는 Mish 함수의 수식을 표현하였으며, Mish 도 Swish 와 유사하게 부드럽고 비 단조함수이다. 양수 영역에서 값이 선형적으로 증가하는 ReLU 와 달리 Mish 는 부드럽게 증가하기 때문에 지속해서 미분이 가능하다. 이는 효과적인 최적화와 일반화를 하는 데 도움이 된다. 그러나 ReLU, Swish 와 비교하여 계산비용이 더 많이 필요해서 학습 시 세대(Epoch) 당 더 많은 시간이 소요된다.

4. 실험

실험에서 학습은 초해상화에서 보편적으로 사용되는 DIV2K[2] 데이터집합을 공통으로 사용하여, EDSR 모델의 활성함수를 ReLU, Leaky ReLU, Swish, Mish 로 변경하면서 각각 학습시켰다. 이때 EDSR 모델은 해상도를 2 배로 변환하는 경우와 4 배로 변환하는 경우를 각각 학습시켰다. 학습을 시킬 때 활성함수 이외의 다른 매개 변수들은 기존의 EDSR 모델과 동일하게 학습을 진행하였다. Leaky ReLU 의 α 값은 0.01 로 설정하여 학습을 진행하였다.

테스트는 EDSR 모델을 초해상화 모델에서 보편적으로 사용되는 총 5 가지 데이터집합(DIV2K[3], SET5[10], SET14[11], B100[12], Urban100[13])을 사용해 진행하였다. 성능 지표로 영상 또는 동영상 손실 압축에서 화질 손실 정보를 평가할 때 사용되는 PSNR 과 영상 품질을 측정하기 위한 구조적 유사도 지수를 구할 때 사용되는 SSIM 을 사용하였다.

<표 1>은 5 가지의 데이터집합 별로 해상도를 2 배로 변환하는 경우와 해상도를 4 배로 변환하는 경우의 활성함수에 따른 성능을 비교한 결과이다. EDSR 에서 기존에 활성함수로 사용하던 ReLU 를 기준으로 하였다.

결과를 보면 해상도를 2 배로 변환하는 경우, 기존 EDSR 모델의 활성함수로 사용되던 ReLU 가 실험에 사용된 다른 활성함수들 보다 비슷하거나 높은 성능을 보인다. ReLU 를 기준으로 PSNR 과 SSIM 이 평균적으로, Leaky ReLU 는 8.50%, 3.58% 낮았고, Swish 는 0.23%, 0.06% 낮았으며, Mish 는 0.23%, 0.07% 낮았다. 종합적으로는 ReLU 가 비교에 사용된 다른 활성함수들보다 PSNR 과 SSIM 에서 평균적으로 약 2.99%, 1.24% 정도 성능이 높은 것을 알 수 있었다.

해상도를 4 배로 변환하는 경우에서는 Leaky ReLU 와 Swish 가 기존에 사용되던 ReLU 보다 약간 더 높은 성능을 보인다. ReLU 를 기준으로 PSNR 과 SSIM 이 평균적으로, Leaky ReLU 는 0.06%, 0.05% 높았고, Swish 는 0.06%, 0.03% 높았으며, Mish 는 0.14%, 0.17% 낮았다.

결과적으로 해상도를 2 배로 변환하는 경우, ReLU 가 가장 성능이 좋았던 결과를 보였으며, ReLU 의 음수 영역을 0 으로 만드는 특성이 다른 초해상화 모델보다 상대적으로 층수가 깊고 채널이 많은 EDSR 모델에 규제효과로 작용하여 가장 좋은 성능을 보인 것으로 해석된다. 이에 비해 해상도를 4 배로 변환하는 경우는 해상도가 4 배로 커지는 만큼 저해상도 영상과 변환된 고해상도 영상 간의 사상(Mapping) 관계가 복잡하다. 따라서 Leaky ReLU 와 Swish 가 ReLU 보다 약간 향상된 성능을 보인 결과는 해상도를 4 배로 변환하는 경우에 사상 관계의 복잡성에 대응하는 모델의 용량이 상대적으로 적절해져 음수 영역을 0 으로 만드는 ReLU 의 규제효과가 큰 의미를 가지지 못한 것으로 판단된다. 이러한 이유로 Leaky ReLU 와 Swish 가 ReLU 보다 성능이 다소 향상된 것으로 해석된다. <표 1>에서 볼 수 있듯이 음수 영역을 0 으로

만들지 않고 값을 어느 정도 유지하는 특성을 가진 Leaky ReLU가 해상도를 2배로 변환하는 경우, ReLU에 대비해 성능이 크게 하락하였지만, 해상도를 4배로 변환하는 경우에서는 오히려 ReLU보다 성능이 다소 향상된 결과가 이러한 해석을 뒷받침한다.

5. 결론

본 논문에서는 초해상화 과업에서 좋은 성능을 보인 EDSR 모델과 ReLU를 대체하기 위해 제안된 활성함수인 Leaky ReLU, Swish, Mish를 비교분석 하였다. 실제 5가지 데이터집합을 사용하여 각 활성함수에 따른 EDSR 모델의 성능을 정량화 하였다. 해상도를 2배로 변환하는 경우, 기존에 사용하던 ReLU가 가장 성능이 좋았지만, 해상도를 4배로 변환하는 경우에서는 Leaky ReLU와 Swish가 ReLU보다 다소 향상된 성능을 보임을 확인하였다.

본 연구에서 해상도를 4배로 변환하는 EDSR 모델에서 Leaky ReLU와 Swish가 ReLU에 대비해 다소 향상된 성능을 보였기 때문에, 향후 연구에서는 ReLU를 활성함수로 사용하는 다른 초해상화 모델에 활성함수를 Leaky ReLU나 Swish로 대체하는 비교실험을 수행하는 것도 필요하다고 판단된다.

사사

이 성과는 2020 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원과 과학기술정보통신부 및 정보통신기획 평가원의 SW 중심대학지원사업으로 수행된 연구임 (No. NRF2018R1C1B5086441)

참고문헌

[1] Saeed Anwar, Salman Khan, and Nick Barnes, "A Deep Journey into Super-resolution: A Survey", arXiv:1904.07523, 1, 2019.

[2] B. Lim, S. Son, H. Kim, S. Nah, and K.M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", In CVPR Workshops, 2017, 1.

[3] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L.Zhang, et al, "Ntire 2017 challenge on single image super-resolution: Methods and results", In CVPR Workshops, 2017, 1,2,4,6,7,8.

[4] C. Ledig, L. Thesis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al, "Photo-realistic single image super-resolution using a generative adversarial network, arXiv:1609.040802, 1,2,3,4,5,6,7, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In CVPR, 2016, 3.

[6] Nair, Vinod and Hinton, Geoffrey E, "Rectified linear units improve restricted Boltzmann machines", In ICML, 2010, pp. 807-814.

[7] Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li, "Empirical Evaluation of Rectified Activations in Convolutional Network", arXiv:1505.00853, 1, 2015.

[8] Prajit Ramachandran, Barret Zoph, and Quoc V. Le, "Swish: a self-gated activation function", arXiv:1710.05941 7, 1, 2017.

[9] Diganta Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function", arXiv:1908.08681, 1, 2019.

[10] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberti-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding", In BMVC, 2012, 2,4.

[11] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations", In Proceedings of the International Conference on Curves and Surfaces, 2010, 2,4.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics.", In ICCV, 2001, 4.

[13] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars.", In CVPR, 2015, 2,4,6.