

생성적 적대 신경망 기반의 딥 러닝 비디오 초 해상화 모델 경량화 및 최적화 기법 연구

김동휘 이수진 *박상호

경북대학교

paka96@knu.ac.kr sujinnie@knu.ac.kr *s.park@knu.ac.kr

A Study on Lightweight and Optimizing with Generative Adversarial Network Based Video Super-resolution Model

Dong-hwi Kim Su-jin Lee *Sang-hyo Park

Kyungpook National University

요약

FHD 이상을 넘어선 UHD급의 고해상도 동영상 콘텐츠의 수요 및 공급이 증가함에 따라 전반적인 산업 영역에서 네트워크 자원을 효율적으로 이용하여 동영상 콘텐츠를 제공하는 데에 관심을 두게 되었다. 기존 방법을 통한 bi-cubic, bi-linear interpolation 등의 방법은 딥 러닝 기반의 모델에 비교적 인풋 이미지의 특징을 잘 잡아내지 못하는 결과를 나타내었다. 딥 러닝 기반의 초 해상화 기술의 경우 기존 방법과 비교 시 연산을 위해 더 많은 자원을 필요로 하므로, 이러한 사용 조건에 따라 본 논문은 초 해상화가 가능한 딥 러닝 모델을 경량화 기법을 사용하여 기존에 사용된 모델보다 비교적 적은 자원을 효율적으로 사용할 수 있도록 연구 개발하는 데 목적을 두었다. 연구방법으로는 structure pruning을 이용하여 모델 자체의 구조를 경량화 하였고, 학습을 진행해야 하는 파라미터를 줄여 하드웨어 자원을 줄이는 연구를 진행했다. 또한, Residual Network의 개수를 줄여가며 PSNR, LPIPS, tOF 등의 결과를 비교했다.

1. 서론

전반적인 OTT 서비스 및 스트리밍 서비스 시장에서의 고해상도의 수요가 높아지고, 시장에서는 8K 해상도의 전자기기까지 상용화되어, 고해상도의 콘텐츠를 쉽게 접할 수 있게 되었다. 더욱이 소비자의 수요는 고해상도의 이미지 혹은 비디오에 점차 집중되었으며 이를 서비스하기 위한 리소스의 부하를 줄이기 위해 부단히 노력을 하고 있는데, 그 중 저해상도의 소스만 가지고 고해상도의 서비스를 제공할 수 있는 초해상화 기술은 CNN, GAN, RNN 등 여러 형태의 딥 러닝 구조를 통해 상용화시키기 위해 노력하고 있다.

그 중 GAN 기반의 이미지 초 해상화 [1]는 기존에 있던 이미지를 작은 해상도의 이미지로 만들어 원본의 고 해상도로 만들기 위해 학습을 진행하였고, 이전의 CNN 구조를 통해 학습된 모델로부터 생성된 결과보다 PSNR 등의 평가결과는 기존보다 떨어지나, 이미지의 더욱 특징을 잘 잡아낸다는 결과를 확인할 수 있었다. 이러한 딥 러닝 기반의 방법을 통한 초해상화 방법은 고 해상도의 이미지를 기본 of bicubic, bilinear 등의 보간법 보다는 조금 더 잘 생성해 낼 수 있다는 장점은 있지만, 비교적 더욱 많은 하드웨어 리소스를 요구를 한다. 이러한 요구사항을 기반으로, 딥 러닝 경량화 기법을 사용하여 비교적 적은 리소스를 가지고 도 모델을 사용할 수 있는 환경을 만드는 것이 본 논문의 목적이다.

2. 경량화 연구방법

딥 러닝 경량화 기법들은 주로 양자화(Quantization) [2, 3], 프루닝(Pruning) [3, 4, 5, 6], 지식 증류(Knowledge-distillation) [6], 가중치 공유(Weight sharing) [7] 등의 방법이 주로 딥 러닝의 경량화 기법으로 알려져 있다. 여러 방법 중 프루닝을 이용하였는데, 프루닝의 경우 크게는 두 방법으로 나눌 수 있다. 학습 후의 중요도가 낮은 가중치, 혹은 랜덤으로 가중치를 제거하는 Unstructure pruning과 구조 자체를 변경시키는 structure pruning 방법이 존재하고, 본 논문의 경우, structure pruning 기법을 사용하였다.

또한 기존에 제시되어 있는 프루닝 [6]의 경우 training, pruning, fine-tuning의 순서를 통해 기존의 정확도를 잃지 않도록 하는데, 이러한 기존의 방법을 그대로 도입하여 연구를 진행하였다.

GAN 기반의 고 해상도의 이미지를 생성해 낼 수 있는 초 해상화 기법으로는 TecoGAN [9] 이라는 생성적 적대 신경망 기반의 모델을 사용하였으며, 사용된 모델은 이미지를 생성해내기 위한 Generator와 실제 이미지와 가짜 이미지를 판별해내는 discriminator의 구조로 되어 있다. 깊게는 TecoGAN 모델의 Generator는 영상 내에 존재하는 Optical flow를 계산하기 위한 FRNet [11], 그리고, reconstruction과 upsampling을 담당하는 SRNet으로 이루어져 있다. 그 중 SRNet에 사용되는 Residual Network의 개수를 적절히 줄여가며 Generator 구조

자체를 변경시키는 Structure pruning을 적용시켜 성능의 하락폭, 구조 자체가 학습 결과에 영향을 미치는 영향성을 파악할 수 있었다.

3. 연구과정 및 결과

TecoGAN의 레이어 구조를 경감시키며 실험을 진행할 수 있었고, super-resolution 네트워크로 사용되는 SRNet은 존재하는 Residual Block의 수를 9개, 7개, 5개, 3개, 1개 순으로 제거해보며 모델의 구성을 변경했다.

표 1은 모델의 학습, 검증, 평가를 위한 데이터 셋을 보여주고 있다. 훈련 데이터 셋은 REDS [12] 데이터 셋을 이용했고, 학습을 위해 만들어야 하는 저해상도의 데이터는 gaussian blurring을 이용한 4배 downsampling으로 생성을 했다.

표 2는 제안된 방법의 학습 환경을 나타낸다. 각각 300,000 iteration의 학습을 진행하며, 테스트는 ToS3, Vid4 데이터 셋을 이용하여 진행했고, 모델에 대한 성능은 PSNR, LPIPS, tOF으로 비교를 하였다. 그리고 각각의 모델을 학습하는 데에 NVIDIA Tesla V100 환경을 기준으로 약 43h 소요되었다.

| | |
|---------------|--------------------|
| Train Dataset | REDS train dataset |
| Valid Dataset | REDS valid dataset |
| Test Dataset | ToS3, Vid4 |

표 1. 데이터 셋

| | |
|---------|--------------|
| OS | Ubuntu 18.04 |
| Python | 3.7.13 |
| PyTorch | 1.11.0 |
| CUDA | 10.2 |

표 2. 학습 환경

| | Params (M) | PSNR (↑) | LPIPS (↓) | tOF (↓) |
|---------------------------------------|------------|----------|-----------|---------|
| Vanilla TecoGAN | 2.5 | 30.75 | 0.0901 | 0.2677 |
| Pruning model #1 (Residual network 7) | 2.2 | 30.34 | 0.0836 | 0.2605 |
| Pruning model #2 (Residual network 5) | 1.9 | 29.88 | 0.0984 | 0.3460 |
| Pruning Model #2 (Residual network 3) | 1.6 | 29.85 | 0.1130 | 0.3843 |
| Pruning Model #3 (Residual network 1) | 1.3 | 29.78 | 0.1020 | 0.3434 |

표 3. 프루닝 결과

표 3은 Vanilla TecoGAN과 프루닝 기법을 적용하여 모델별로 Parameter 수, PSNR, LPIPS, tOF의 최대 성능을 보여주고 있다. 프루닝 결과, Pruning Model #3은 Vanilla TecoGAN과 비교하여 Parameter 수가 48% 감소했지만 PSNR은 1dB 감소, LPIPS는 0.0119 증가, tOF는 0.0757 증가함을 확인할 수 있다. 즉, 신경망 구조를 작게 만들었음에도 정량적으로 성능이 크게 차이가 나지 않는 것을 확인할 수 있다.

4. 결론

본 논문은 여러 딥 러닝 경량화 방법들 중 Structure Pruning을 통해 실제로 구조를 변경해 보고, 변경된 구조가 얼마나 영향을 끼치는지 확인할 수 있었고, 신경망 구조를 비교적 작게 만들었음에도 불구하고, 정량평가에서는 값의 결과가 떨어지지 않음을 확인할 수 있었다. 하지만 결과 이미지에 대해서는 PSNR과 LPIPS만을 이용해 결과를 비교했기 때문에, 정량평가에 관한 결과와는 다르게 실제 사람이 느끼는 바와는 다를 수 있다고 느껴질 수 있을 것이라 예상된다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구 지원과(2021-0-01082), 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00087, SD/HD급 저화질 미디어의 고품질 변환 기술 개발)

참고 문헌 (References)

- [1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
- [2] Anwar, S., Hwang, K., & Sung, W. (2015, April). Fixed point optimization of deep convolutional neural networks for object recognition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1131-1135). IEEE.
- [3] Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *arXiv preprint arXiv:1510.00149*.
- [4] LeCun, Y., Denker, J., & Solla, S. (1989). Optimal brain damage. *Advances in neural information processing systems*, 2.
- [5] Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.

- [6] Frankle, J., & Carbin, M. (2018). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*.
- [7] Liu, Z., Sun, M., Zhou, T., Huang, G., & Darrell, T. (2018). Rethinking the value of network pruning. *arXiv preprint arXiv:1810.05270*.
- [8] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 27).
- [9] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [10] Chu, M., Xie, Y., Leal-Taixé, L., & Thurey, N. (2018). Temporally coherent gans for video super-resolution (tecogan). *arXiv preprint arXiv:1811.09393*, 1(2), 3.
- [11] Sajjadi, M. S., Vemulapalli, R., & Brown, M. (2018). Frame-recurrent video super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6626-6634).
- [12] Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., & Mu Lee, K. (2019). Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).