

## 모바일 단말에서 실시간으로 동작하는 초고해상화 기술 개발

\*김성제 \*\*정진우 \*\*\*GANZORIG GANKHUYAG

한국전자기술연구원

\*sungjei.kim@keti.re.kr

## Real-time Single Image Super Resolution in Mobile Devices

\*Kim, Sungjei \*\*Jeong, Jinwoo

Korea Electronics Technology Institute

### 요약

본 논문은 모바일 단말에서 실시간으로 동작하는 딥러닝 기반 경량 초고해상화 기술에 관한 내용이다. 대용량 3차원 메쉬 모델의 비실시간 압축은 실시간 스트리밍 응용 시나리오에서 제약점으로 작용하고 있고, 본 논문에서는 두 입력 텐서의 차원을 일치시켜야 하는 element-wise 덧셈 대신 concatenation을 활용해 연산량을 개선하고, float-to-int8 양자화 과정에서 발생하는 오차를 줄이기 위해 weight clipping 및 regularization 기법을 활용해 초고해상화 화질 성능을 개선하였다. 제안하는 알고리즘은 기존 모바일 초고해상화 기술을 화질 측면에서 0.12dB, 처리 속도 측면에서 13.6ms 개선하였고, Mobile AI & AIM 2022 실시간 이미지 초고해상화 대회에서 1등을 수상하였다.

### 1. 서론

단일 영상 기반 초고해상화 기술은 저해상도 영상을 입력으로 하여 고해상도 영상을 생성하는 기술로, 원격 감지 및 수중 영상처리, 의료 영상처리, 멀티미디어 응용 분야 등 컴퓨터 비전 응용 분야에서 널리 사용되고 있다. 최근에는 딥러닝 기술을 활용한 초고해상화 기술들[1]의 연구가 많이 진행되었고, 전통적인 보간 기법 기반 고해상화 기술[2]의 성능을 압도적으로 개선하였다.

하지만, 대부분의 단일 영상 초고해상화 기술 연구는 고성능 GPU에서 처리가 가능한 네트워크에 대한 것으로, IoT나 모바일 단말과 같은 컴퓨팅 자원에 제약이 있는 환경에서 실시간으로 동작하는 딥러닝 네트워크에 관한 연구[3]는 상대적으로 부족한 상황이다.

본 논문에서는 모바일 단말에서 높은 성능과 실시간 처리 속도를 갖는 경량 초고해상화 네트워크 기술에 대해서 제안한다. 실험 결과를 통해 제안하는 알고리즘이 DIV2K 테스트셋에 대해서 처리속도 기준으로 19.20 ms를 달성하면서도 PSNR 기준으로 30.03dB를 달성하였음을 보였다.

### 2. Anchor-based Plain Net

기존 초고해상화 네트워크는 Convolution Neural Network (CNN) 레이어를 연속적으로 쌓아서 성능을 향상하게 시키기에는 제약이 있어 skip-connection 레이어를 네트워크 중간에 넣어 vanishing gradient를 완화하여 네트워크의 학습 성능을 개선했다[4]. 하지만 feature map 도메인에서 skip-connection은 float-to-int8 변환 과정에서 양자화 에러가 커지는 문제가 있어 이를 해결하기 위해서 Anchor-based Plain Net (ABPN)에서는 입력 영상을 복제해서 네트

워크 최종단에서 element-wise 덧셈 연산으로 병합하는 방법을 제안하였다[5]. 하지만, ABPN 방법은 element-wise 덧셈 연산을 위해 두 입력 차원을 갖게 처리하는 과정에서 입력 영상을 여러 번 복제하게 되었고, 이로 인해 많은 처리 시간을 소비하게 되었다.

### 3. 제안하는 초고해상화 네트워크

본 논문에서는 입력 영상을 여러 번 복제하는 과정에서 발생하는 메모리 복사 연산을 최소화하기 위해서 입력 영상을 바로 feature map 도메인에서 연결해서 합치는 Skip-concatenated 네트워크를 제안하였다. 그림 1과 같이 Skip-concatenated 네트워크는 8비트 입력 영상을 양자화 없이 바로 feature map과 합칠 수 있기 때문에, float-to-int8 변환 과정에서 발생하는 양자화 에러가 최소화되는 장점이 있고 처리 속도가 빠르다는 장점이 있다.

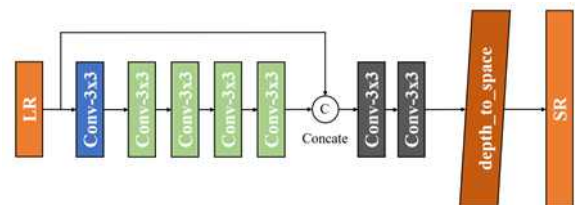


그림 1 Skip-Concatenated 초고해상화 네트워크

또한, float-to-int8 양자화 과정에서 에러를 최소화하기 위해 양자화 인식 학습 기법인 (Quantization-Aware Training, QAT) 방법을 도입하였고, QAT 과정에서 에러를 최소화하기 위해서 weight clipping 및 regularization 기법을 도입하여 학습된 weight의 분포가 최대한

symmetry 형태를 유지할 수 있도록 제한을 하였다. 해당 제한 기법을 통해서 QAT 이후에도 초기 학습 성능에서 0.1dB 하락 수준으로 유지할 수 있었다.

#### 4. 실험 결과

제안 알고리즘의 성능 평가를 위해 DIV2K 데이터셋의 800장은 학습에 활용하고, 100장은 검증에 활용하였다. 표 1에서는 제안 기법에 따른 PSNR의 측정 결과이다. Scratch는 weight를 랜덤 초기화한 이후, 학습한 방법을 의미하며, QAT는 Scratch를 통해 학습된 weight를 초기값으로 하여 학습한 결과이다. TFLite(int8)은 QAT 결과 weight를 int8로 변환하는 과정이다. 표 1은 언급한 각 과정이 끝난 이후 측정된 PSNR 결과값이다. SCSRN\_w/o\_WC는 weight clipping을 적용하지 않은 결과이고, SCSRN\_w/\_WC는 weight clipping을 적용한 결과이다. 전체 양자화 과정에서 (Scratch -> TFLite(int8)) weight clipping을 적용했을 때, 약 0.095dB 정도 하락하였고, 적용하지 않을 때 비해서 개선된 것을 확인할 수 있었다.

<표 1> 제안하는 네트워크의 PSNR 측정 결과

PSNR(dB)	Scratch	QAT	TFLite (int8)
SCSRN_w/o_WC	<b>30.328</b>	<b>30.242</b>	<b>30.195</b>
SCSRN_w/_WC	<b>30.323</b>	<b>30.230</b>	<b>30.228</b>

#### 5. 결론

본 논문에서는 모바일 단말에 적합한 Skip-concatenated 네트워크를 제안하고, 양자화 과정에서 발생하는 오차를 최소화하기 위해서 weight clipping 및 regularization 기법을 제안하였다. 제안하는 기술은 기존 모바일 초고해상화 기술을 화질 측면에서 0.12dB, 처리 속도 측면에서 13.6ms 개선하였고, ECCV에서 진행된 Mobile AI & AIM 2022 실시간 이미지 초고해상화 대회에서 1등을 수상하였다.

#### Acknowledgements

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00802, 속성을 유지하는 지능적 미디어 화면비변환 기술 개발)

#### 참고문헌

- [1] K. Hayat, "Multimedia super-resolution via deep learning: A survey," *Digital Signal Processing*, 81:198-217, 2018.
- [2] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE transactions on Image Processing*, 15(8):2226-2238, 2006.
- [3] A. Ignatov, et al, "Real-Time Quantized Image Super Resolution on Mobile NPUs, Mobile AI 2021 Challenge: Report," In *Proceedings of the IEEE conference on computer vision and*

*pattern recognition workshops*, pp. 2525-2534, 2021.

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136-144, 2017.

[5] Z. Du, J. Liu, J. Tang, and G. Wu, "Anchor-based plain net for mobile image super-resolution," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2494-2502, 2021