

주파수 대역 분할 및 그룹 컨볼루션을 이용한 개선된 고속 푸리에 컨볼루션

허재호, 배주한, 주현철, 나태영
에스케이텔레콤

{jaeho.hur, juhan.bae, hc.joo, taeyoung.na}@sk.com

Improved Fast Fourier Convolution Using Frequency Band Division and Grouped Convolution

Jaeho Hur, Juhan Bae, Hyunchul Joo, Taeyoung Na
SK Telecom

요약

본 논문에서는 기존 Fast Fourier Convolution(FFC)의 모델링 역량을 개선하기 위한 새로운 접근법을 제안한다. 기존 FFC는 주파수 도메인에서 모든 주파수 성분에 동일한 1×1 Convolution 가중치를 적용하는 한계를 가지고 있어, 주파수 성분별로 차별화된 학습이 불가능하다. 본 논문에서는 입력 Feature를 균등한 크기의 여러 하위 주파수 대역으로 분할하고, 각 주파수 대역마다 독립적인 가중치를 적용할 수 있는 Grouped Convolution 기반의 새로운 FFC 구조를 제안한다.

I. 서론

Deep Neural Network에서 Convolutional Neural Network의 Receptive Field를 효과적으로 확장하여 모델링 역량을 높이기 위해 Fast Fourier Convolution(FFC)[1]이 제안되었다. FFC는 Fast Fourier Transform을 기반으로 주파수 도메인에서 하나의 Feature 값을 변경하여 Spatial Domain의 입력 Feature 전체에 영향을 미칠 수 있는 성질을 이용한다. 이를 통해 보통의 Convolution과 달리 Receptive Field의 범위를 전역으로 넓히는 것이 가능하며, Local Receptive Field와 Non-local Receptive Field의 수용 및 처리를 하나의 Unit에서 동시에 가능하도록 하는 효과적인 앙상블 구조를 제안하였다.

그러나 기존 FFC의 핵심 구성 요소인 Fourier Unit(FU)에서는 주파수 도메인에 적용되는 1×1 Convolution의 가중치가 모든 주파수 성분에 동일하게 적용되기 때문에, 주파수 성분마다 다른 Convolution 가중치로 학습하는 것이 불가능한 한계가 존재한다.

본 논문에서는 이러한 기존 FFC의 제한된 모델링 역량을 개선하기 위해 주파수 대역 분할 및 Grouped Convolution[2] 기반의 새로운 FFC 구조를 제안한다.

II. 제안하는 방법

이번 장에서는 기존 FFC의 핵심 연산 단위인 FU에 대해 간략히 설명하고, 제안하는 방법이 주파수 대역 분할과 Grouped Convolution을 이용하여 기존 FU의 한계점을 어떻게 개선하는지 설명한다.

(1) 기존 FU 구조

FU는 먼저 Fast Fourier Transform(이하 FFT)을 통해 입력 Feature를 주파수 도메인으로 변환하고, 주파수 도메인에서 1×1 Convolution, Batch Normalization, ReLU를 통해 일반적인 Deep Neural Network의 연산을 수행한다. 마지막 과정에서 Inverse FFT를 통해 다시 Spatial Domain으로 Feature를 복원하여 마무리한다. 여기에서 주파수 도메인에 적용되는 1×1 Convolution의 가중치는 모든 주파수

성분에 동일하게 적용되기 때문에, 주파수 성분마다 다른 Convolution 가중치로 학습하는 것이 불가능한 한계가 있다. 그림 1의 (a)는 기존 FU의 구조를 나타낸다.

(2) 제안하는 개선된 FU 구조

제안하는 FU는 기존 FU와 마찬가지로 FFT를 통해 입력 Feature를 주파수 도메인으로 변환하는 과정부터 시작한다. 이후, Frequency Splitter, 1×1 K-Grouped Convolution, Frequency Band Merger와 같은 새로운 구성 요소들을 거쳐 최종적으로 Inverse FFT를 통해 Feature를 복원한다. 이러한 구조는 연산 효율성을 극대화하면서도 각 주파수 대역별로 정교한 모델링을 가능하게 하여 전체 모델의 표현력을 향상시킨다. 그림 1의 (b)는 제안하는 개선된 FU 구조를 나타내며, 세부 동작에 대한 설명은 다음과 같다.

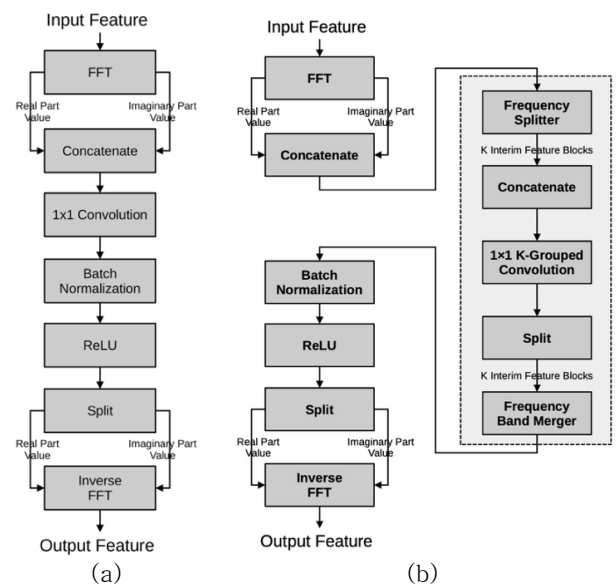


그림 1. 기존의 FU와 제안하는 개선된 FU 구조: (a) 기존 FU 구조, (b) 주파수 분할 및 Grouped Convolution 기반의 개선된 FU 구조

Frequency Splitter: FFT 를 통해 얻어진 주파수 도메인의 Feature Tensor 를 균등한 크기의 여러 개의 하위 주파수 대역으로 분할한다. 분할의 기준은 텐서의 공간 차원(Height, Width) 상에서 $N \times M$ 크기로 정확하게 K 개로 나누는 것이다. 이 과정을 통해 K 개의 Interim Feature Block 들이 생성되며, 이는 주파수 도메인에서의 저주파 및 고주파 성분들을 물리적으로 분리하는 효과를 가져온다.

Concatenate: 분할된 K 개의 Interim Feature Block 들은 첫 번째 채널 축을 기준으로 연결(Concatenate)되어, 1×1 K-Grouped Convolution 의 입력으로 사용할 수 있는 단일한 Tensor 형태로 변환된다.

1×1 K-Grouped Convolution: 연결된 Feature Tensor 는 K 개의 그룹을 가진 1×1 Grouped Convolution 레이어에 입력된다. 이 레이어는 각 그룹(즉, 각 주파수 대역)에 대해 독립적인 1×1 Convolution 을 수행하여, 각 주파수 대역의 특성에 맞는 개별적인 가중치를 학습하게 된다. 이 단계는 제안하는 방법의 핵심으로, 기존 FU 의 단일 1×1 Convolution 이 가지는 모델링 역량의 한계를 극복하는 역할을 한다.

Split/Frequency Band Merger: 1×1 K-Grouped Convolution 의 출력은 다시 첫 번째 채널 축을 기준으로 분리(Split)되어 개별 Frequency Band 에 대응하는 Interim Feature Block 들로 복원된다. 이후, Frequency Band Merger 를 통해 분리된 주파수 대역들이 원래의 공간 차원 위치로 재배열 및 병합된다. 이 과정은 Grouped Convolution 으로 처리된 정보가 기존의 FFC 구조에 다시 원활하게 통합되도록 보장한다.

제안하는 새로운 FU 는 기존 FFC 의 FU 를 완전히 대체할 수 있으며, 전체 FFC 네트워크 구조의 큰 변경 없이도 모델의 표현력을 획기적으로 향상시킬 수 있는 이점을 가진다. Grouped Convolution 의 채택은 구현의 용이성과 병렬 연산 효율성 극대화 측면에서도 긍정적인 효과를 가져올 수 있다.

III. 실험 결과

제안하는 방법의 효과를 객관적으로 검증하기 위해, Image Inpainting Task 를 실험 대상으로 선정하였다. Image Inpainting 은 이미지의 손상되거나 가려진 부분을 주변 맥락에 맞게 자연스럽게 복원하는 작업으로, 전역적인 이미지 구조와 질감을 이해하는 능력이 요구된다. 따라서 FFC 와 같은 전역적 정보 처리 모델의 성능을 평가하기에 매우 적합한 Task 이다.

성능 비교를 위해, 기존 FFC 기반 Image Inpainting 모델인 Large Mask Inpainting[3] 을 기준으로 설정하였다. 제안하는 모델은 이 기준 모델의 FU 를 새롭게 개선한 FU 로 교체하여 구성하였다. 모델의 성능은 FID, LPIPS, SSIM 등의 세 가지 주요 평가 지표를 사용하여 측정하였다.

표 1. Image Inpainting Task 성능 비교

평가 지표	기존 FFC 기반 Inpainting Model	제안하는 FFC 기반 Inpainting Model
FID (↓)	42.94	41.514
LPIPS (↓)	0.126	0.121
SSIM (↑)	0.871	0.873

표 1에서 나타내는 바와 같이, 실험 결과는 FID, LPIPS, SSIM 모든 지표에서 새로운 FFC 기반 모델이 기존

모델을 능가하는 성능을 보였다. 특히, FID 가 42.94 에서 41.514 로 감소한 것은 매우 중요한 결과이다. FID 는 생성된 이미지의 전체적인 품질과 사실성을 종합적으로 평가하는 지표로, 이 수치의 개선은 제안하는 FU 구조가 기존 구조보다 더 풍부한 특징을 학습하고, 실제 이미지의 데이터 분포에 더 가까운 고품질의 이미지를 생성할 수 있게 되었음을 의미한다. 이는 본 논문이 해결하고자 했던 기존 FFC 의 제한된 모델링 역량 문제를 성공적으로 해결했음을 입증한다.

또한, LPIPS 와 SSIM 의 개선은 생성된 이미지가 원본 이미지와 지각적 및 구조적으로 더 유사하다는 것을 의미한다. 이는 복원된 영역이 주변부와 더 자연스럽게 어우러지고, 이미지의 핵심적인 구조적 정보가 손상 없이 잘 복원되었음을 나타낸다. 이러한 결과는 주파수 대역별로 특화된 학습을 통해 모델이 저주파 성분(전역적 구조)과 고주파 성분(국소적 디테일)을 모두 효과적으로 처리할 수 있게 되었기 때문이다. Grouped Convolution 을 주파수 대역 분할과 결합한 것은 단순한 연산 효율성 향상을 넘어, FFC 의 표현력을 근본적으로 향상시키는 구조적 효과를 가져왔다.

IV. 결론

본 논문은 기존 FFC 의 제한된 모델링 역량 문제를 해결하기 위한 새로운 방법을 제시하였다. 제안하는 방법은 기존 FFC 의 핵심 유닛인 FU 의 구조를 개선하기 위해 주파수 대역 분할 및 Grouped Convolution 을 도입하여, 주파수별 특성을 독립적으로 학습할 수 있는 메커니즘을 구축했다. 이를 통해 Grouped Convolution 이 가진 연산 효율성 이점을 활용하면서도, 주파수 도메인에서의 정교한 모델링을 통해 FFC 의 표현력을 성공적으로 향상시킬 수 있음을 입증하였다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광연구개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536, 기여율: 100%)

참 고 문 헌

- [1] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 4479–4488. Curran Associates, Inc., 2020.
- [2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
- [3] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pages 3172–3182, 2022.