

효율적인 CSI 피드백을 위한 합성곱-트랜스포머 융합 모델

허성욱, 고영채*

고려대학교

huhbaksa@korea.ac.kr, *koyc@korea.ac.kr

A Convolution-Transformer Fusion Model for Efficient CSI Feedback

Sunguk Huh, Young-chai Ko
Korea Univ.

요약

FDD 기반 대규모 MIMO 시스템에서는 채널 상호성이 보장되지 않아, 다운링크 빔포밍을 위해 사용자 단말(UE)이 채널 상태 정보(CSI)를 기지국(BS)으로 피드백해야 하며, 이로 인해 막대한 오버헤드가 발생한다. 본 논문에서는 지역적·전역적 특성을 동시에 학습하는 ConvTransNet을 제안한다. 제안 모델은 합성곱 신경망(CNN)과 트랜스포머(Transformer)를 결합하여 낮은 연산 복잡도로도 높은 복원 성능을 달성한다. COST2100 데이터셋 실험 결과, 합성곱 신경망으로 구성된 CsiNet+ 대비 0.8M 낮은 연산량으로 유사하거나 더 나은 NMSE 성능을 보였으며, 트랜스포머로 구성된 TransNet 대비 약 33% 낮은 연산량을 유지하면서도 경쟁력 있는 NMSE 성능을 보였다.

I. 서론

FDD 기반 MIMO 시스템에서는 다운링크와 업링크 채널 간 상호성이 보장되지 않아, 다운링크 빔포밍을 위해 사용자 단말이 기지국으로 CSI를 피드백해야 하며, 이로 인해 특히 대규모 안테나 환경에서 막대한 오버헤드가 발생한다. 이러한 문제를 해결하기 위해 CSI 행렬의 효율적인 압축 및 복원 기법에 대한 연구가 활발히 진행되어 왔다. 2018년 제안된 CsiNet[1]은 최초의 딥러닝 기반 모델로 합성곱 신경망을 활용하여 기존 압축 센싱 방법보다 높은 복원 성능을 보였으며, 이후 CsiNet+[2]은 합성곱 신경망 구조를 개선하여 성능을 향상시켰다. 2022년 제안된 TransNet[3]은 [4]에서 소개된 트랜스포머 구조를 활용해 CSI 피드백 품질을 크게 개선하며 당대 최고 성능을 달성하였다. 본 연구에서는 CSI 행렬이 갖는 전역적 특성과 지역적 특성을 동시에 고려하여, 합성곱 블록(convolutional block)으로 지역 정보를 추출하고 트랜스포머로 전역 정보를 학습하는 ConvTransNet을 제안한다. 제안 모델은 낮은 연산 오버헤드로 우수한 복원 성능을 달성함을 확인하였다.

II. 본론

A. 시스템 모델 및 문제 정의

본 논문에서는 [1]을 따라, 다수의 송신 안테나($N_t \gg 1$)를 갖는 단일 셀 다운링크 대규모 MIMO 시스템을 고려한다. 시스템은 OFDM을 기반으로 하며, 각 부반송파에 서의 수신 신호는 다음과 같이 표현된다:

$$\mathbf{y}_n = \tilde{\mathbf{h}}_n^H \mathbf{v}_n \mathbf{x}_n + \mathbf{z}_n \quad (1)$$

여기서 $\tilde{\mathbf{h}}_n, \mathbf{v}_n, \mathbf{x}_n, \mathbf{z}_n$ 는 각각 채널 벡터, 프리코딩 벡터, 송신 심볼, 잡음을 나타낸다. 모든 부반송파의 채널 벡터

를 결합하면 전체 CSI 행렬 $\tilde{\mathbf{H}}$ 를 얻을 수 있으며, UE는 파일럿 신호를 기반으로 이를 추정하여 BS로 피드백한다. 그러나 송신 안테나 및 부반송파의 개수가 증가할수록 CSI 차원이 급격히 커지며, 이로 인해 피드백 오버헤드가 크게 증가하는 문제가 발생한다.

이를 해결하기 위해 본 연구에서는 2차원 DFT를 적용하여 각도-지연(angular-delay) 도메인으로 변환하고, 유효한 N_a 개의 행만 선택하여 축소된 CSI 행렬 \mathbf{H}_a 를 구성한다. 이후 인코더 f_{en} 을 통해 \mathbf{H}_a 를 저차원 잡재 표현 \mathbf{s} 로 변환하고, 디코더 f_{de} 를 이용하여 이를 복원한다.

$$\mathbf{s} = f_{en}(\mathbf{H}_a, \theta_{en}) \quad (2)$$

$$\hat{\mathbf{H}}_a = f_{de}(\mathbf{s}, \theta_{de}) \quad (3)$$

최종적으로, 본 과정은 평균제곱오차(MSE)를 최소화하는 최적화 문제로 귀결되며, 그 목표는 다음과 같다:

$$(\hat{\theta}_{en}, \hat{\theta}_{de}) = \underset{\theta_{en}, \theta_{de}}{\operatorname{argmin}} \| \mathbf{H}_a - \hat{\mathbf{H}}_a \|^2 \quad (4)$$

B. 제안하는 모델 ConvTransNet

그림 1은 제안하는 ConvTransNet의 구조이다. 입력 CSI 행렬 \mathbf{H}_a 는 복소수 행렬로, 크기는 $N_a \times N_t$ 이다. 이 두 행렬은 결합되어 convolutional block의 입력으로 사용되며, 최종 입력은 $2 \times N_a \times N_t$ 크기의 실수 행렬로 표현된다. convolutional block의 출력은 $2N_a N_t$ 크기의 1차원 벡터로 변환되며, 이후 완전연결층(Fully Connected layer)을 거쳐 N_{fb} 차원으로 압축된다. 이 벡터는 코드워드(codeword) \mathbf{s} 의 형태로 CSI 디코더에 전달된다.

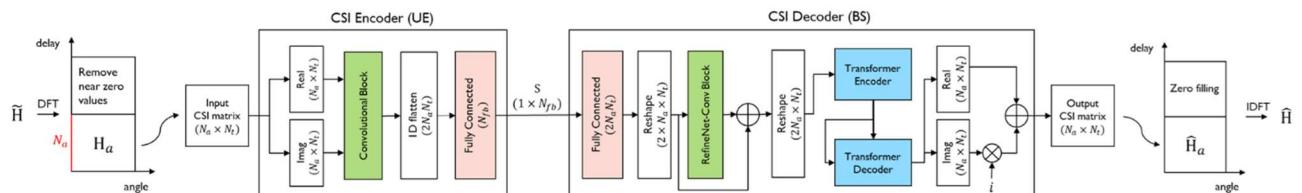


그림 1. ConvTransNet의 구조

표 1. 제안하는 모델과 기존 모델들의 NMSE(dB) 및 FLOPs 성능 비교

Methods	1/4			1/8			1/16			1/32			1/64		
	FLOPs	NMSE(dB)		FLOPs	NMSE(dB)		FLOPs	NMSE(dB)		FLOPs	NMSE(dB)		FLOPs	NMSE(dB)	
		Indoor	Outdoor												
ConvTransNet (proposed)	23.77M	-31.86	-12.02	22.72M	-19.84	-7.75	22.19M	-13.17	-5.69	21.93M	-8.82	-3.61	21.8M	-5.94	-2.69
TransNet	35.72M	-32.38	-14.86	34.70M	-22.91	-9.99	34.14M	-15.00	-7.82	33.88M	-10.49	-4.13	33.75M	-6.08	-2.62
CsiNet+	24.57M	-27.37	-12.40	23.52M	-18.29	-8.72	23.00M	-14.14	-5.73	22.74M	-10.43	-3.40	22.61M	/	/
CsiNet	5.41M	-17.36	-8.75	4.37M	-12.70	-7.61	3.84M	-8.65	-4.51	3.58M	-6.24	-2.81	3.45M	-5.84	-1.93

CSI 디코더는 입력된 코드워드를 완전연결층을 통해 $2N_aN_t$ 크기로 복원한 후, 이를 $2 \times N_a \times N_t$ 형태로 형상 변환하여 RefineNet-Conv block에 전달한다. RefineNet-Conv block에서는 잔차 연결(residual connection)이 적용되며, 출력은 다시 $2 \times N_a \times N_t$ 크기로 형상 변환되어 트랜스포머에 입력된다. 마지막으로 트랜스포머 인코더와 디코더를 거친 출력은 실수부와 허수부 행렬로 분리된다.

ConvTransNet 인코더에서 Convolutional block은 CsiNet[1]의 인코더 구조를 기반으로 하며, 이를 통해 UE의 연산 복잡도를 최소화하면서도 지역적 특성을 효과적으로 추출할 수 있도록 설계하였다. ConvTransNet 디코더에서 RefineNet-Conv block은 지역적 패턴을 포착하고, 트랜스포머는 전역적 상관관계를 학습하여 복원 성능을 향상시킨다. 또한, RefineNet-Conv block에는 잔차 연결을 추가하여 안정성과 성능을 더욱 강화하였다.

C. 실험 환경 및 평가 지표

본 연구에서는 [1]의 설정을 기반으로, COST2100 데이터셋[5]을 사용하여 실험을 수행하였다. 고려한 시나리오는 5.3 GHz 대역의 실내 피코셀 환경과 300 MHz 대역의 옥외 농촌 환경의 두 가지이다. 기지국 안테나 수 (N_t)는 32개로 구성하였으며, FDD 시스템에서 서브캐리어 수(N_c)는 1024, 유효 딜레이 도메인 성분 수 (N_a)는 32로 설정하였다. 트랜스포머 인코더 및 디코더의 차원 d_{model} 은 32로 설정하였다.

데이터셋은 총 150,000 개의 샘플로 이루어져 있으며, 이 중 100,000개는 학습용, 30,000개는 검증용, 20,000개는 테스트용으로 활용하였다. 학습 과정에서는 배치 크기 200과 학습률 1×10^{-4} 를 적용하였고, 드롭아웃은 사용하지 않았다. 다양한 압축 비율(1/4, 1/8, 1/16, 1/32, 1/64)에 대해 400epoch 동안 훈련을 진행하고 NMSE를 평가하였다.

성능 지표로는 원본 CSI 행렬 H 와 복원된 CSI 행렬 \hat{H} 간의 차이를 나타내는 정규화 평균제곱오차(NMSE)를 사용하였으며, 이는 다음과 같이 정의된다:

$$NMSE = E \left\{ \frac{\|H - \hat{H}\|_F^2}{\|H\|_F^2} \right\} \quad (5)$$

D. 실험 결과

표 1은 ConvTransNet과 기존 모델들의 NMSE 성능 및 FLOPs를 비교한 결과를 제시한다. 각 압축 비율(η)에 따른 NMSE 값은 실내 및 실외 시나리오에서 측정되었다. 전반적으로 가장 우수한 NMSE 성능을 보이는 모델은 TransNet이지만, 동시에 가장 높은 FLOPs를 요구한다. 이에 비해 ConvTransNet은 TransNet 대비 약 33% 낮은 FLOPs를 가지면서도 NMSE 성능은 경쟁력 있는 수준을 유지한다. 또한, ConvTransNet의 FLOPs는 CsiNet+ 보다 약 0.8M 작으며, 두 모델 간 NMSE 성능은 시나리오와 압축률에 따라 우위가 달라지지만, 전체적으로 유사한 수준을 보인다. 즉, 특정 조건에서는 CsiNet+가 더 우수하고, 다른 조건에서는 ConvTransNet이 더 나은 결과를 보인다. 결과적으로,

표 1은 ConvTransNet이 합성곱 신경망과 트랜스포머 구조를 결합함으로써 최소한의 연산 복잡도 증가로도 NMSE 성능을 상당히 향상시킬 수 있음을 보여준다.

III. 결론

본 논문에서는 FDD 대규모 MIMO 시스템을 위한 새로운 CSI 피드백 프레임워크 ConvTransNet을 제안하였다. 이 모델은 경량 합성곱 신경망 인코더로 UE의 복잡도를 줄이고, RefineNet-Conv 블록과 트랜스포머를 결합해 국소·전역 채널 특징을 모두 활용한다. 분석 결과, ConvTransNet은 기존 모델 대비 UE 연산량(FLOPs)을 최소화하면서도 COST2100 데이터셋에서 NMSE 성능을 우수하게 달성하였다. 이는 합성곱 신경망과 트랜스포머의 결합이 차세대 무선통신에서 효율적이고 정확한 CSI 피드백을 제공할 수 있음을 보여준다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2024-00428780, 6G-클라우드 리더십구축을 위한 교육 연구 오픈 허브)

참 고 문 헌

- [1] C. -K. Wen, W. -T. Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," in IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [2] J. Guo, C. -K. Wen, S. Jin and G. Y. Li, "Convolutional Neural Network-Based Multiple-Rate Compressive Sensing for Massive MIMO CSI Feedback: Design, Simulation, and Analysis," in IEEE Transactions on Wireless Communications, vol. 19, no. 4, pp. 2827–2840, April 2020.
- [3] Y. Cui, A. Guo and C. Song, "TransNet: Full Attention Network for CSI Feedback in FDD Massive MIMO System," in IEEE Wireless Communications Letters, vol. 11, no. 5, pp. 903–907, May 2022.
- [4] A. Vaswani et al., "Attention is all you need," in Proc. 31st Int. Conf. Neural Inf. Process. Syst. (ICONIP), vol. 30, 2017, pp. 5998–6008.
- [5] L. Liu et al., "The COST 2100 MIMO channel model," in IEEE Wireless Communications, vol. 19, no. 6, pp. 92–99, December 2012.