

# 다중 심도에서 추출한 특징 벡터의 계층적 융합을 이용한 딥 JSCC 시스템

김혁, 채찬병\*

연세대학교 글로벌융합공학과

{hyukkim9933, cbchae}@yonsei.ac.kr

## Deep JSCC with Multi-Depth Feature Extraction and Hierarchical Fusion

Hyuk Kim, Chan-Byoung Chae\*

Yonsei Univ.

### 요약

본 논문은 이미지 전송의 효율성과 복원 성능을 향상시키기 위해 다중 심도 특징 추출(Multi-Depth Feature Extraction)과 계층적 융합(Hierarchical Fusion)을 적용한 새로운 Deep Joint Source-Channel Coding(DepJSCC) 구조를 제안한다. 기존의 딥 JSCC 시스템은 주로 단일 경로의 특징 추출에 의존하여, 다운샘플링에 따른 국소적 정보 손실을 줄이기 위해 다운샘플링 단계마다 특징의 dimension 을 증가시켰다. 다운샘플링 단계마다 특징을 추출하여 트랜스포머 블록을 덜 거친 분기를 생성하는 제안된 방식은 파라미터 수를 크게 절약하면서도 전역적·국부적 정보를 효과적으로 결합한다. 시뮬레이션 결과, 기존 딥 JSCC 대비 동등하거나 더 우수한 복원 성능을 확인하였다.

### I. 서론

무선 이미지 전송에서 소스 부호화와 채널 부호화를 분리하는 전통적인 Separation Scheme 은 이론적으로 최적이지만, 실제 환경에서는 채널 상태 악화 시 Cliff Effect 가 발생하여 복원이 불가능해진다<sup>[1]</sup>. 이를 해결하기 위해 최근에는 소스와 채널을 결합해 학습하는 Deep Joint Source-Channel Coding(DepJSCC)이 주목받고 있다.<sup>[3]</sup> DepJSCC 는 채널 상태 변화에 따라 성능이 점진적으로 저하되는 Graceful Degradation 특성을 가지며, 기존 통신 시스템에 융화시키는 연구도 이루어지고 있다.<sup>[2]</sup>

초기 Deep JSCC 는 주로 CNN 기반 백본을 사용했으나, CNN 은 국소적 특징 추출에는 강점이 있지만 전역적 문맥 정보를 처리하는 데 한계가 있다. 이를 보완하기 위해 최근에는 Vision Transformer, 특히 Swin Transformer<sup>[4]</sup>를 활용한 구조가 제안되었다. Swin Transformer 는 윈도우 기반 계층적 구조를 통해 전역적·국부적 정보를 모두 효과적으로 처리하며, 이미지 복원 성능에서 CNN 기반 모델을 능가하는 결과를 보여준다. 대표적인 예로 SwinJSCC<sup>[6]</sup>와 NTSCC<sup>[5]</sup>는 Swin Transformer 기반 비선형 변환을 적용해 기존 DepJSCC 대비 우수한 성능을 달성하였다.

그러나 Transformer 는 Convolution 연산에 매우 비싸다. 특히 Swin Transformer 의 경우, 다운샘플링에 따른 해상도 저하로 인한 정보 손실을 메우기 위해

다운샘플링이 진행될 때마다 dimension 을 2 배씩 늘리는데<sup>[4]</sup>, 이는 트랜스포머의 연산을 기하급수적으로 증가시킨다.

또한, 최근 연구에서는 단일 경로의 특징 추출을 넘어 Multi-Stage Feature Extraction 을 적용하는 시도가 이루어지고 있다<sup>[7]</sup>. 이는 다양한 심도에서 추출된 특징을 융합해 더 풍부한 표현을 제공하며, JSCC 에도 적용 가능하다. 다중 심도 특징을 활용하면 다운샘플링 과정에서 발생하는 정보 손실을 줄이고 복원 성능을 향상시킬 수 있다.

본 논문에서는 이러한 배경을 바탕으로, 다중 심도 특징 추출과 계층적 융합을 결합한 새로운 Deep JSCC 구조를 제안한다. 제안 방식은 파라미터 수를 크게 절약하면서도 전역적·국부적 정보를 효과적으로 결합하여 이미지 복원 성능을 극대화한다.

### II. 본론

본 논문에서는 이미지 전송의 효율성과 복원 화질을 극대화하기 위해 다중 심도 특징 추출(Multi-Depth Feature Extraction) 및 계층적 융합(Hierarchical Fusion) 기법을 적용한 새로운 딥 소스-채널 결합 부호화(DepJSCC) 구조를 제안한다. 기존의 딥 JSCC 시스템은 주로 단일 경로의 특징 추출에 의존하여, 다운샘플링에 따른 국소적 정보 손실을 줄이기 위해 다운샘플링 단계마다 특징의 dimension 을 증가시켰다.

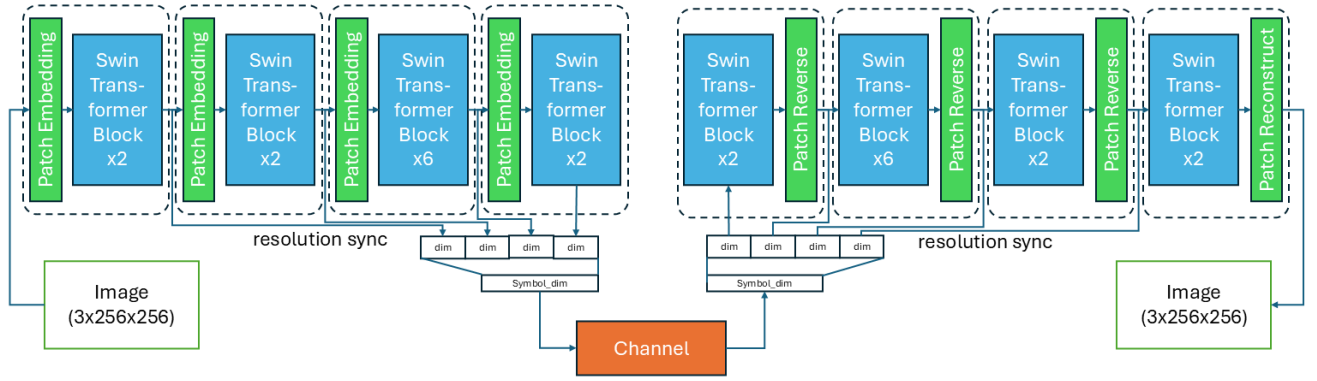


그림 1. 전체 구조도

제한하는 구조의 인코더는 다운샘플링 단계마다 분기(branch)를 생성하여, 트랜스포머 레이어를 거치는 횟수가 각기 다른 다단계 특징 벡터들을 추출한다. 추출된 특징들은 모두 다운샘플링 레이어를 통과했기에 동일한 공간적 크기(Spatial dimension)를 유지한 채널 단위로 결합(Concatenation)되어 전송 심볼로 투영된다.

디코더에서는 수신된 심볼을 다시 다단계 특징으로 분리하고, 이를 복원 과정의 적절한 위치에 재삽입하여 계층적으로 융합한다. 특히 디코더의 트랜스포머 레이어는 재삽입된 특징과 이전 단계의 특징을 동시에 처리할 수 있도록 설계되어, 전역적(Global) 정보와 국부적(Local) 정보를 효과적으로 결합한다. 시뮬레이션 결과, 제안된 기법은 차원을 늘리지 않아 패러미터 수를 적게 유지함에도 기준에 차원을 늘린 딥 JSCC 와 동일한 수준의 우수한 이미지 복원 성능을 보임을 확인하였다.

### III. 결론

본 논문에서는 다중 심도에서의 feature 추출과 재결합으로 기존의 Swin Transformer 기반 딥 JSCC 시스템에서 패러미터와 연산량을 각각 10 배, 3 배 절약하면서도 성능을 비슷한 수준으로 유지하였다.

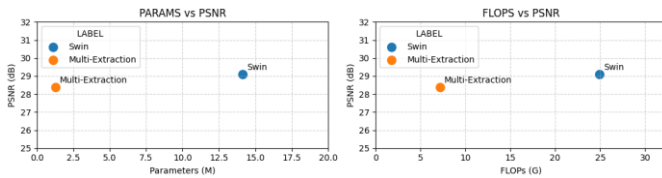


그림 2. 성능 비교

그러나 패러미터 및 연산량 감소에도 불구하고 학습 시간은 크게 단축시킬 수 없었다. 이는 Layer Norm, Softmax 및 Swin Transformer 의 구조적 특징 (매번 window 분할 및 shifting) 등이 부동소수점 연산(FLOPs)는 거의 차지하지 않으나 시간이 많이 걸리기 때문이라고 사료된다. 따라서 후속연구로는 backbone 을 Swin Transformer 에서 Dilated Neighborhood Attention 으로 교체하고, 한 트랜스포머 내에서 다양한 수준의 dilation 을 가하여 receptive field 를 늘리는 방향으로 실험할 예정이다.

### ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2024-00428780, 6G-Cloud Research and Education Open Hub),

This work was supported by the National Research Foundation of Korea (NRF) Grant through the Ministry of Science and ICT (MSIT), Korea Government, under Grant 2022R1A5A1027646

### 참 고 문 헌

- [1] E. Boursoulatz, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," IEEE Transactions on Cognitive Communications and Networking
- [2] M. Yang, C. Bian, and H. S. Kim, "Deep joint source channel coding for wireless image transmission with OFDM," Proc. ICC 2021-IEEE International Conference on Communications
- [3] D. Gündüz, M. A. Wigger, T. Y. Tung, P. Zhang, and Y. Xiao, "Joint source-channel coding: Fundamentals and recent progress in practical designs," Proceedings of the IEEE.
- [4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," Proc. IEEE/CVF International Conference on Computer Vision (ICCV),
- [5] J. Dai et al., "Nonlinear transform source-channel coding for semantic communications," IEEE Journal on Selected Areas in Communications,
- [6] K. Yang et al., "Swinjscc: Taming swin transformer for deep joint source-channel coding," IEEE Transactions on Cognitive Communications and Networking,
- [7] W. Shi, J. Xu, and P. Gao, "Sformer: A lightweight transformer for semantic segmentation," Proc. 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSp),