딥러닝 기반 이미지전송 시스템을 위한 Top-K Water-filling 알고리즘

박범수, 이남윤
고려대학교

bumsupark@korea.ac.kr, namyoon@korea.ac.kr

A Study on the ********** ********** systems

Bumsu Park, Namyoon Lee
Korea Univ.

요 약

Transformer 구조를 이용한 이미지 전송연구는 는 $O(N^2d+Nd^2)$의 높은 계산 복잡도와 많은 파라미터를 갖기 때문에, 계산과 메모리 측면에서 비효율적이다. 본 연구에서는, 2-stage top-k feature selection 과 water-filling 을 활용한 새로운 JSCC 방법을 제시한다. 제안하는 방법은 메모리를 요구하지 않으며, $O(N\log N)$의 낮은 복잡도를 갖는다. 실험 결과, 2 중 Transformer 구조 대비 4000 배의 적은 연산만으로 2.5dB 이내의 peak-signal-to-noise-ratio 성능 차이를 달성해 제안하는 방법의 효율성을 확인하였다.

Ⅰ. 서 론

딥 러닝을 이용해 이미지를 전송한 첫 연구 [1]는, 고전 알고리즘보다 좋은 압축-복원 성능을 달성하여, 많은 후속연구가 잇따랐다. 그 중에서도 [2]는 [1]의 Convolutional neural network (CNN)를 성능이 우수한 Transformer 구조로 대체해, 탁월한 성능을 달성하여 주목받았다. 하지만 Transformer 구조는 $O(N^2d+Nd^2)$ 의 높은 계산 복잡도를 갖고, 많은 매개변수를 사용한다. 따라서, 전력과 저장공간이 제한된 통신 상황에선 Transformer 를 적용하기 힘들다는 문제가 있고, 이를 해결하려는 시도 [4]들은 주로 경량화 된 신경망구조를 도입하는 방식을 사용했다. 하지만, 이러한 방식은 신경망 구조를 사용하는 이상 매개변수를 필요로 한다는 한계가 있다. 따라서 본 논문에서는 1) 전체 신경망 구조의 일부를 고전 알고리즘으로 대체하는 Top-K water-filling 방식을 제안하고, 2) 실험을 통해, 제안하는 Top-K water-filling 이 신경망 구조기반의 방법과 동일한 성능을 달성하는 동시에 전체 매개변수 수와 계산 복잡도를 크게 줄이는 것을 확인한다.

Ⅱ. 본론

1. 시스템 모델

본 논문은 그림. 1 과 같은 딥 러닝을 이용한 이미지 전송 시스템을 고려한다. 전체 시스템은 인코더, 채널, 디코더로 이루어져 있다. 인코더는 이미지 $\mathbf{x}\in\mathbb{R}^N$와 $\mathbf{x}$를 입력으로 받아 잠재변수 $\mathbf{y}\in\mathbb{R}^M$로 압축하는 전처리 함수 g와, $\boldsymbol{y}$를 입력 받아 hyperprior $\boldsymbol{z}\in\mathbb{R}^L$를 구하는 함수 h, $\boldsymbol{z}$로부터 $\boldsymbol{y}$의 평균과 분산을 추정하는 $\hat{h}$, 추정한 평균과 분산 $\boldsymbol{\mu},\boldsymbol{\sigma}^2$를 이용해 $\boldsymbol{y}$를 채널입력 $\boldsymbol{s}\in\mathbb{R}^K$로 변환하는 joint source-channel coding(JSCC) encoder 로 이루어져 있다. 채널은 $\boldsymbol{s}$에 노이즈를 추가해 수신단에서 noisy 한 $\hat{\boldsymbol{s}}$를 수신하도록 하며, 이때 채널은 σ_n^2를 분산으로 갖는 AWGN 채널을 가정한다.

디코더는 hyperprior $\boldsymbol{z}$로부터 얻은 $\boldsymbol{\mu},\boldsymbol{\sigma}^2$를 이용해 $\hat{\boldsymbol{s}}$로부터 $\hat{\boldsymbol{y}}$를 추정하는 JSCC decoder $\hat{g}$, $\hat{\boldsymbol{y}}$로부터 복원 이미지 $\hat{\boldsymbol{x}}$를 구하는 후처리 $\hat{h}$로 이루어진다.

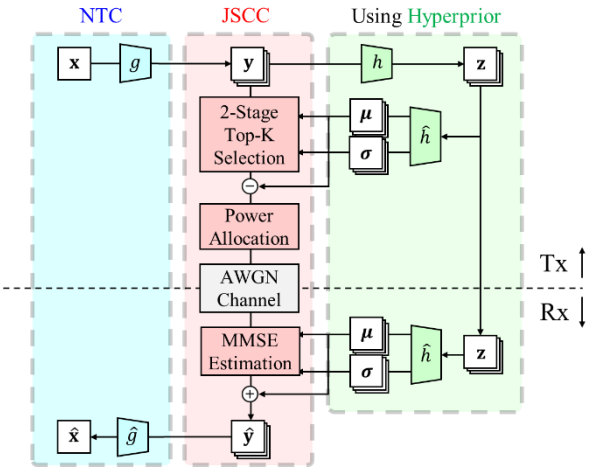

그림 1. 시스템 모델

이때 풀고자 하는 문제와 제약조건은 다음과 같다:

$$\mathbb{E}[\|x-\hat{x}\|_2^2] \tag{1}$$

$$\dim(s)=K \tag{2}$$

우리가 풀려는 문제는, 위 가정을 배경으로, K 번의 channel 을 사용해서 이미지를 전송할 때, 원본과 복원 이미지 간의 오차를 최소화하는 것이다.

2. 신경망 기본구조

본 논문은, $g,h,\hat{g},\hat{h}$ 는 neural network 이고, $\boldsymbol{z}$는 reliable 한 채널을 통해 error 없이 전달된다고 가정한다. 또한, $g,h,\hat{g},\hat{h}$ 는 [2]의 손실함수에 최적화되었다고 가정한다. [2]의 가정에 따르면, $\boldsymbol{y}$의 분포는 다음 가정을 따른다:

$$P_{\boldsymbol{y}} = \prod_{i=1}^{M} P_{y_i}(y_i), \qquad (3)$$

$$P_{y_i|z}(y_i) \sim \mathcal{N}(\mu_i, \sigma_i^2).$$

위 시스템 모델에 따라, 본 논문의 문제는 기존의 이미지 전송 문제에서 latent 를 정확히 보내는 latent 전송 문제로 바뀐다.

3. Top-k & water-filling

본 논문은 시스템 모델링에 기반한 top-k selection, water-filling 솔루션을 제안한다. Top-k selection 은 $\boldsymbol{y}$ 값의 일부를 선택하여, 보내는 정보의 차원을 줄이는 방법으로, 본 논문에선 constraint 1 (dimension constraint)를 만족하는 동시에 objective 를 최소화하는 K 개를 뽑는다. 우리는 selection 기준으로 $\sigma^2, -\log p(y), |y-\mu|$ 를 고려했다. 그 중 |y-mu|, entropy 를 기준으로 뽑을 때 가장 이미지 복원 오차가 작았지만, |y-mu|, entropy 는 데이터로부터 realize(번역 필요)해야 하는 값으로, 디코더에서 이미지를 복원하기 위해선 뽑힌 K 값의 인덱스(번역필요) 정보가 필요하다. 시뮬레이션을 통해 복원오차와 인덱스 오버헤드(번역필요)의 트레이드오프를(번역필요) 비교한 결과, sigma 를 기준으로 뽑을 때, Rate distortion 성능이 가장 좋아 topk 선택 기준으로 사용한다.

뽑힌 K 개의 feature, $y-\mu$ 는, 수식 (1)에 따라 가우시안 분포로 가정된다. rate-distortion 이론[3]에 따르면, 독립 가우시안 소스(번역필요)이 AWGN 채널을 통해 보내질 때, MSE 관점에서 최적 인코딩, 디코딩 방법은 다음과 같은 단순 scalar multiplication 이다:

$$s_i = p_i(y_i - \mu_i),$$

$$\hat{s}_i = \frac{p_i^2\sigma_i^2}{p_i^2\sigma_i^2+1}(y_i-\mu_i) + \frac{p_i^2\sigma_i^2}{p_i^2\sigma_i^2+1}n_i, \qquad (4)$$

이때 p_i는 average power constraint 를 다음과 같이 만족한다:

$$\sum_{i=1}^{K} \sigma_i^2 p_i^2 = P. \qquad (5)$$

주어진 (2), (3)에 따라 MSE 는 다음과 같이 계산되고,

(6)

MSE 를 최소화하는 power allocation 문제는 다음과 같이 정립된다:

(7)

이는 p 에 대한 convex problem 으로, KKT condition 을 이용하면 다음과 같은 solution 을 구할 수 있다.

(8)

4. 시뮬레이션

본 논문은 CLIC[4] dataset 을 압축-전송하는 시뮬레이션을 통해, 제안하는 Topk waterfiiling 과 DeepJSCC[1], NTSCC[2]와의 압축-복원 성능을 비교한다. 시뮬레이션 설정은 다음과 같다, 실험하는 K 의 범위는 {범위}, AWGN 채널 노이즈의 SNR 은 10dB 이고, Top-k water-fiiling 의 backbone model 은 NTSCC 구조와 동일하고 NTSCC 의 손실함수를 최소화하도록 최적화된 모델이다. 평가 지표는 입력 이미지의 차원 대비 채널 사용횟수의 비율인 CBR, 과, CBR 에 대한 PSNR 을 사용한다.

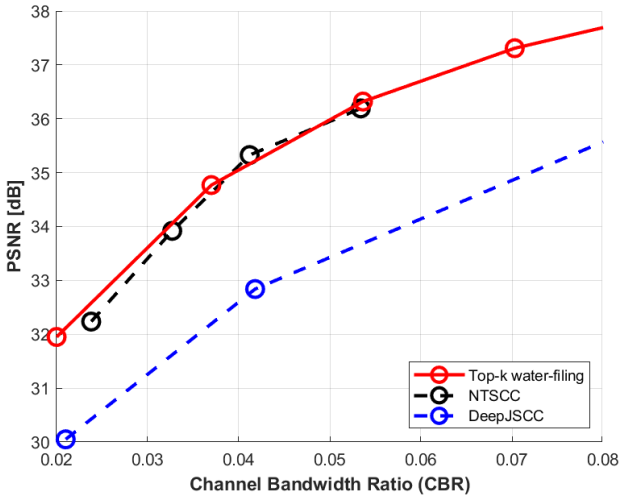

그림 3. 시뮬레이션 CBR-PSNR 성능

(단위: 1,000,000)

	NTSCC	DeepJSCC	Top-K water filling
FLOPs	400,000		106.5
파라미터	41.2/10.3	0.88/0.22	0

표 1. FLOPs, 파라미터 수 비교.

그림 3 의 시뮬레이션 결과에서 중요한 두 가지 사실을 확인할 수 있다. 첫 번째는 CNN 구조를 사용하는 DeepJSCC 대비 Transformer 구조를 사용하는 NTSCC, top-k waterfilling 이, 모든 CBR 영역에서 2dB 이상 PSNR 차이를 보인다는 점에서, Transformer 구조가 CNN 대비 우수 성능을 갖는다는 것이다. 두 번째는, Transformer 를 사용해 latent 를 압축-전송하는 NTSCC 와, 고전 알고리즘을 사용해 latent 를 압축-전송하는 top-k waterfilling 이 같은 성능을 달성한다는 점에서, transformer 로 잘 압축된 latent 는 고전 통신만으로도 잘 전송할 수 있다는 점이다.

또한, 표 1 에서 확인할 수 있듯, Top-K water-filling 은 NTSCC 대비 적은 파라미터를 사용하고, X 배의 연산 량 만을 사용하여, 제안하는 방법의 효율성을 확인할 수 있다.

결론

본 논문은 Transformer 기반 이미지 전송 시스템이, $O(N^2d + Nd^2)$ 의 높은 계산 복잡도를 갖고, 많은 파라미터 수를 사용하는 문제를 해결하고자, Top-K selection, water-filling 솔루션을 제안한다. Top-K selection, water-filling 은 $O(N \log K)$ 의 낮은 계산 복잡도를 가질 뿐만 아니라, JSCC 과정에서 파라미터를 사용하지 않아, JSCC 에 transformer 를 사용하는 NTSCC 대비 적은 파라미터 수를 갖는다. 실험을 통해, Top-K water-filling 이 NTSCC 와 동일한 성능을 달성하는 것을 보이고, 제안하는 Top-K water-filling 의 좋은 성능을 확인한다.

참 고 문 헌

[1] Bourtsoulatze, Eirina, David Burth Kurka, and Deniz Gündüz. "Deep joint source-channel coding for wireless image transmission." *IEEE Transactions on Cognitive Communications and Networking* 5.3 (2019): 567-579.

[2] Dai, Jincheng, et al. "Nonlinear transform source-channel coding for semantic communications." *IEEE*

Journal on Selected Areas in Communications 40.8 (2022): 2300-2316.

[4] Li, Weizhi, et al. "Non-orthogonal multiple access enhanced multi-user semantic communication." IEEE Transactions on Cognitive Communications and Networking (2023).

[3] Cover, Thomas M. *Elements of information theory*. John Wiley & Sons, 1999.

[4] CLIC dataset