Deep Joint Source-Channel Coding and Modulation for Underwater Acoustic Communication

*Yoshiaki Inoue[1], Daisuke Hisano[1],

Kazuki Maruta[2], Yuko Hara-Azumi[2], Yu Nakayama[3]

Osaka University[1]
Tokyo Institute of Technology[2]
Tokyo University of Agriculture and Technology[3]

JSCC

- Ordinary image transmission systems

  - Employ two-step encoding

- Joint Source-Channel Coding (**JSCC**)

  - Images are converted directly to the transmitted bits

Joint Source-Channel Coding and Modulation (**JSCCM**)

- **Encoder** $f_{\text{enc}} : \mathbb{R}^{H\times W\times 3} \to \mathbb{C}^{K\times 1}$
  - Outputs **complex symbols** $\boldsymbol{s}$ from the input **image** $\boldsymbol{x}$
- **Decoder** $f_{\text{dec}} : \mathbb{C}^{K\times 1} \to \mathbb{R}^{H\times W\times 3}$
  - Outputs an **images** $\tilde{\boldsymbol{x}}$ from the input **complex symbols** $\tilde{\boldsymbol{s}}$

Deep JSCCM

JSCCM based on Deep Learning (**Deep JSCCM**)[1]

- Parameterization with deep neural networks (DNNs)
  - $f_{\mathrm{enc},\boldsymbol{\theta}}$: Encoder function, $f_{\mathrm{dec},\boldsymbol{\theta}}$: Decoder function
- The set of parameters $\boldsymbol{\theta}$ are learned via backpropagaction
  - The error between input and output images are minimized
  - Utilize stochastic gradient descent (SGD) type optimizers
- Superiority over conventional methods has been reported[1]

[1] E. Bourtsoulatze et al., in *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, 2019.

Deep JSCCM

JSCCM based on Deep Learning (**Deep JSCCM**)[1]

- Parameterization with deep neural networks (DNNs)
  - $f_{\mathrm{enc},\boldsymbol{\theta}}$: Encoder function, $f_{\mathrm{dec},\boldsymbol{\theta}}$: Decoder function
- The set of parameters $\boldsymbol{\theta}$ are learned via backpropagaction
  - The error between input and output images are minimized
  - Utilize stochastic gradient descent (SGD) type optimizers
- Superiority over conventional methods has been reported[1]
- However, previous works assume simple theoretical channels:
  - AWGN and Slow Reyleigh fading channels

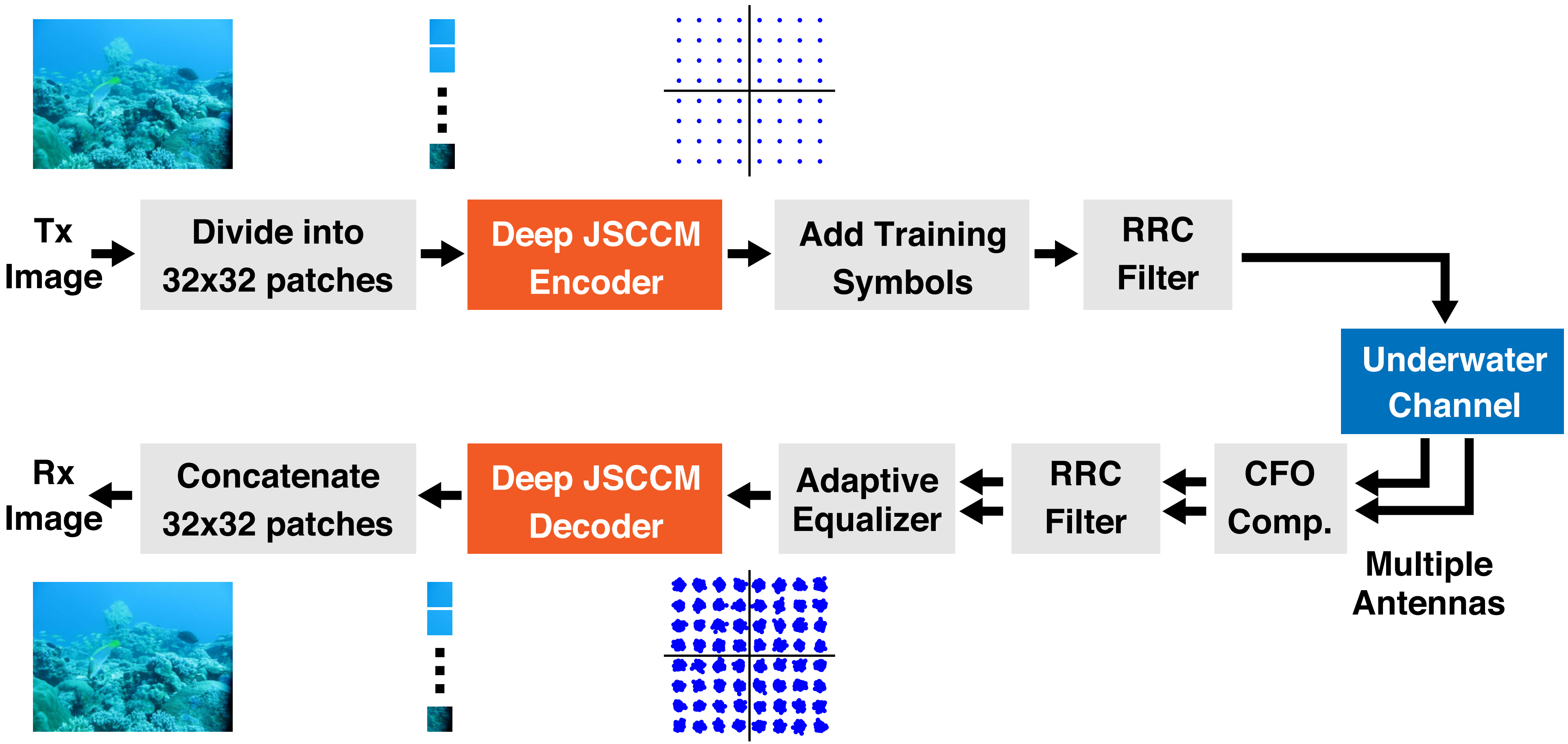
Overview of This Work

- We extend the deep JSCCM technique to
  underwater acoustic image transmission
- Underwater image transmission is a fairly challenging task
  - Limited transmission rate due to narrow bandwidth
  - Severe intersymbol inference (ISI) by long-delay multipaths
- We present a new deep JSCCM approach incorporating
  - Long-delay multipath communication channel
  - Practical transmitter and receiver filters
  - Time-domain adaptive equalizer (TD-AE) to mitigate ISI
- Effectiveness of the proposed scheme is shown with simulation

System Model and Proposed Method

System Overview

- The system overview is shown below
  - Single-Input-to-Multi-Output (SIMO) channel



Channel Model

$\mathbf{x} \in \mathbb{C}^{1\times N_\mathrm{s}}$: Transmission signal vector (N_s: # of samples)

$\mathbf{y}_i \in \mathbb{C}^{1\times N_\mathrm{s}}$: Reception signal vector (ith antenna), $\mathbf{Y}$: Matrix of $(\mathbf{y}_i)$

- We consider the following linear AWGN channel model

$$\mathbf{Y} = \mathbf{H}\mathbf{x} + \mathbf{Z}$$

$\mathbf{H}$: Channel matrix, $\mathbf{Z}$: AWGN matrix

Channel Model

$\mathbf{x} \in \mathbb{C}^{1 \times N_s}$: Transmission signal vector (N_s: # of samples)

$\mathbf{y}_i \in \mathbb{C}^{1 \times N_s}$: Reception signal vector (ith antenna), $\mathbf{Y}$: Matrix of $(\mathbf{y}_i)$

- We consider the following linear AWGN channel model

$$\mathbf{Y} = \mathbf{H}\mathbf{x} + \mathbf{Z}$$

$\mathbf{H}$: Channel matrix, $\mathbf{Z}$: AWGN matrix

- $\mathbf{H}$ consists of block matrices $\mathbf{H}_i$ (i: Antenna index)[2]

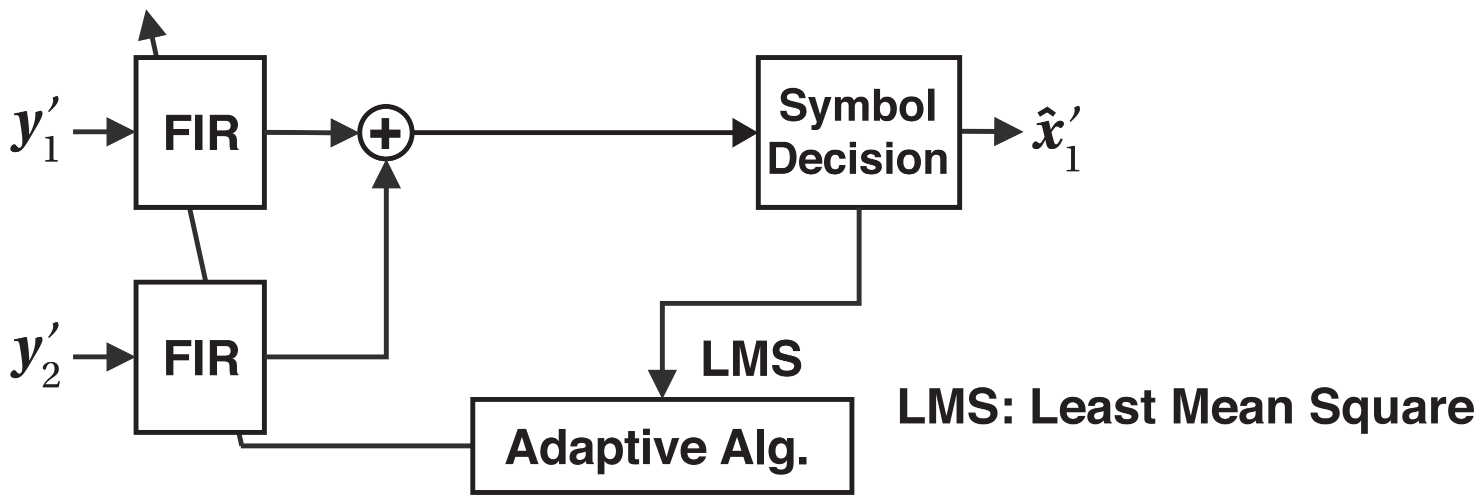
$$h_i^{(k,l)} = \sqrt{L_i^{(k,l)}} \exp\left[j(2\pi f_i^{(k,l)} k T_s + \phi_i^{(k,l)}) \right]$$

Diag. elements: Direct wave, Non-diag. elements: Delayed wave

$L_i^{(k,l)}$: Relative power of delayed wave, $f_i^{(k,l)} := f_d \cos \alpha_i^{(k,l)}$, f_d: Doppler freq.
T_s: Sampling freq., $\phi_i^{(k,l)}$: Initial phase, $\alpha_i^{(k,l)}$: Antenna incident angle

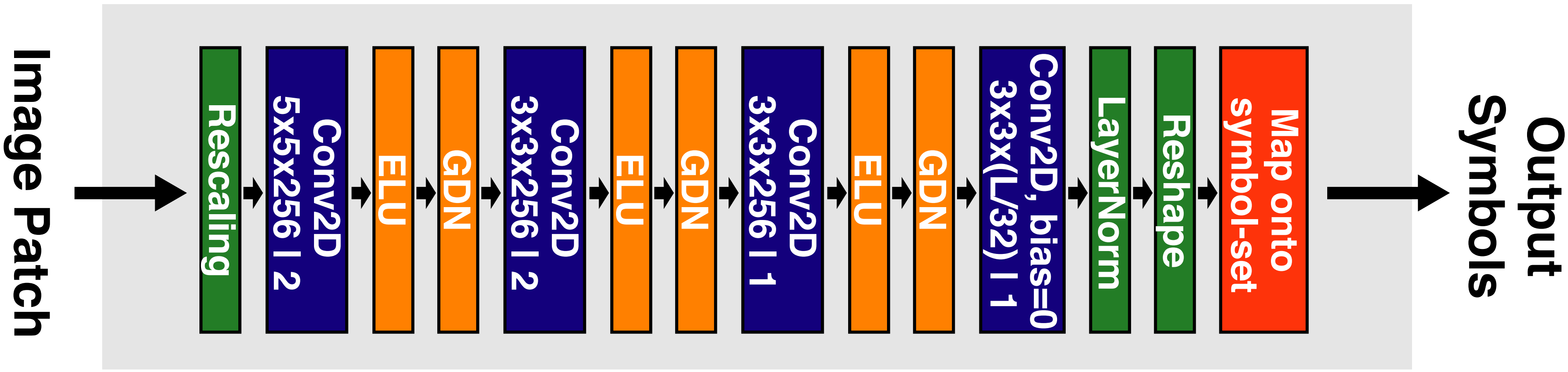
[2] K. Shima et al., in *IEEE Access*, vol. 9, pp. 18361–18372, 2021.

Time-Domain Adaptive Equalizer (TD-AE)



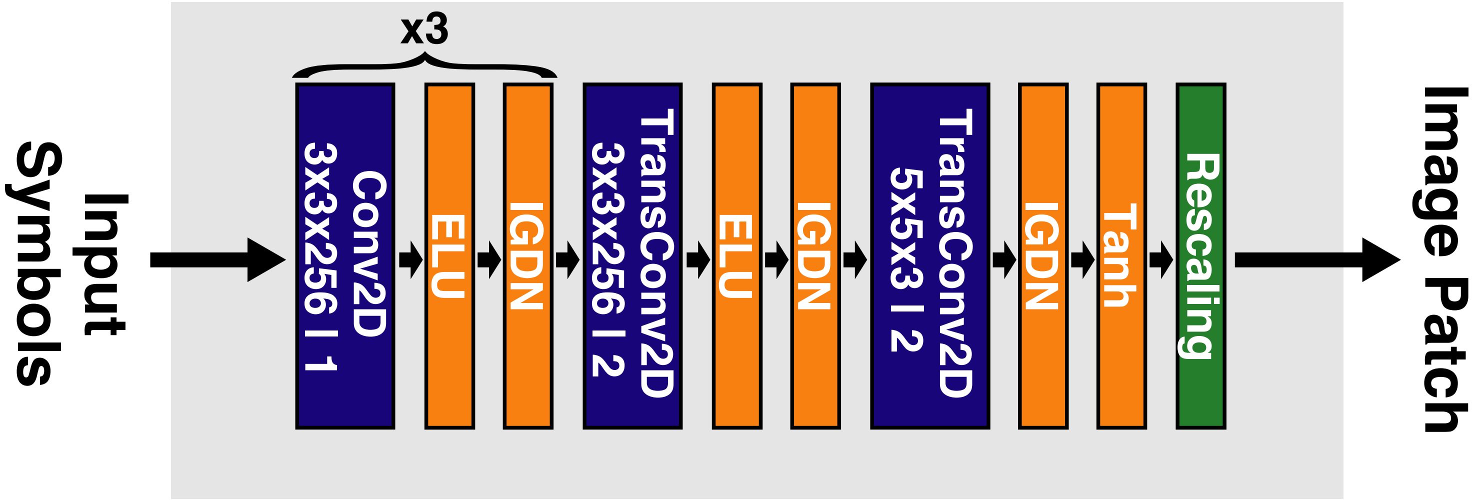
- Finite impulse response (FIR) filter exists for each antenna
  - FIR filters' outputs are combined
- The coefficients of the FIR filters are updated successively based on symbol-decision feedback

JSCCM Encoder



- Conv2D: Convolution layer (Notation: kernel_shape | strides)
- ELU: Exponential Linear Unit
- GDN: Generalized Divisive Normalization
- Map onto symbol-set ★Necessary to employ the TD-AE
  - Outputs the nearest point on a (trained) symbol set $S \subset \mathbb{C}^M$ (M: # of candidate symbols)

JSCCM Decoder



- TransConv2D: Transposed convolution layer (Notation: kernel_shape | strides)
- IGDN: Inverse Generalized Divisive Normalization

Training method (1)

- Non-differentiable functions in the system
  - Transmitter and receiver filters
  - Linear AWGN communication channel
  - TD-AE of the receiver

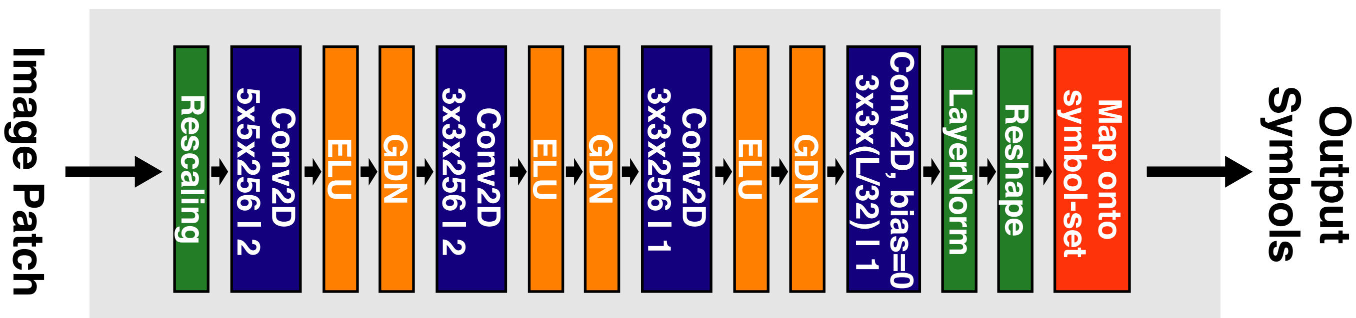
- For a non-differentiable function ψ,
  the following procedure enables backpropagation
  - Consider the relation $\boldsymbol{y} = \psi(\boldsymbol{x})$
  - Replace this relation by the following equation:

$$\boldsymbol{y} = \boldsymbol{x} + \mathrm{sg}(\psi(\boldsymbol{x}) - \boldsymbol{x})$$

sg: Stop gradient operator (the input is regarded as a constant)

Training method (2)

- We employ the Adamax Optimizer[3]
- Loss function: Mean squared error (MSE) of received image
  - Exception: Symbolset S in the encoder's last layer



- Training of the symbol set S
  - Jointly trained with other layers (using the Adamax optimizer)
  - Loss function: MSE of the layer's input and output (in S)

[3] D. Kingma and J. Ba, in *Proc. ICLR 2015*, 2015.

Training Method (3)

- Our model contains complicated non-differentiable layers
  - Long-delay multipath communication channel
  - Practical transmitter and receiver filters
  - Time-domain adaptive equalizer (TD-AE) to mitigate ISI

➨ It is difficult to train the encoder/decoder from sctatch

- Two-step training procedure to overcome this difficulty
  - **Pre-Training**: Employs a simple AWGN channel model
    - ➨ Suitable initial values for the parameters are obtained
  - **Main-Training**: Employs the targeted system model

Numerical Simulation

Simulation Parameters

- We employed a two-path channel model

For comparison, we employed

- Image compression:
  - JPEG, JPEG2000
- Modulation:
  - QPSK, 16QAM
- Forward error correction:
  - Turbo code (1/2, 3/4)

(Channel parameters are taken from [4])

Item	Value
# of antennas	2
Carrier freq.	300 kHz
System bandwidth	200 kHz
# of training symbols	4000
Roll-off rate of RRC filters	0.2
# of FIR taps in TD-AE	21
SNR	18–26 dB
Relative moving speed	1 m/s
Relative delay of delay wave	520 samples
Relative power ratio $L_i^{(k,l)}$	0.4
Speed of sound	1500 m/s
Incident angle $\alpha_i^{(k,l)}$	Random
Initial phase $\phi_i^{(k,l)}$	Random

[4] H. Fukumoto et al., in *Proc. Global Oceans 2020*, 2020.

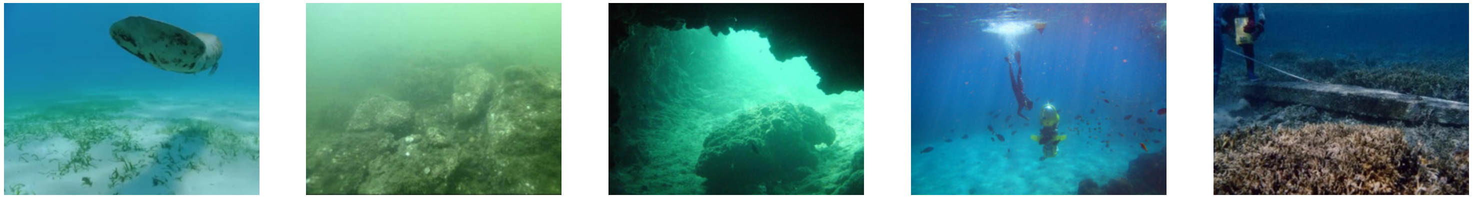
Datasets

- Training dataset : Downsampled Imagenet[5]
  - Consists of many 32×32 color images (1,281,167 for training and 50,000 for validation)

[5] P. Chrabaszcz et al., arXiv preprint, arXiv:1707.08819, 2017.

Datasets

- Training dataset : Downsampled Imagenet[5]
  - Consists of many 32×32 color images (1,281,167 for training and 50,000 for validation)
- Test dataset : Underwater image dataset[6] (890 images)



[5] P. Chrabaszcz et al., arXiv preprint, arXiv:1707.08819, 2017.
[6] C. Li et al., in *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

Datasets

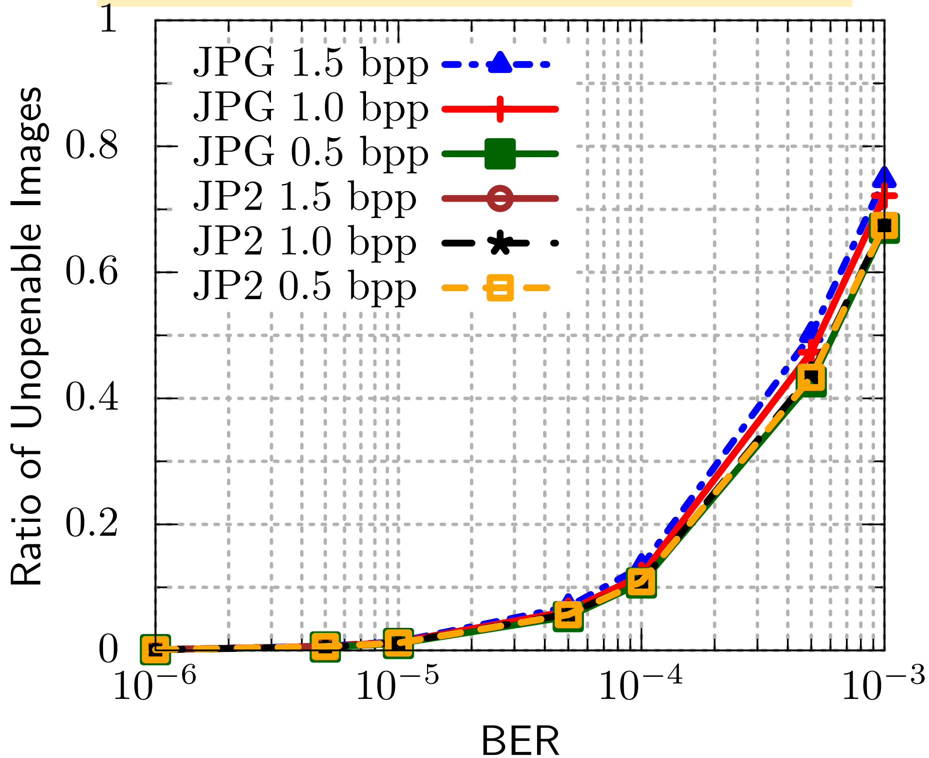
- Training dataset : Downsampled Imagenet[5]
  - Consists of many 32×32 color images (1,281,167 for training and 50,000 for validation)
- Test dataset : Underwater image dataset[6] (890 images)
  - Preprocessed to obtain equal-size images
    - Extract images with $(\text{width}/\text{height}) \geq 4/3$ (825 images)
    - Resize and crop to obtain images with 256×192 pixels

[5] P. Chrabaszcz et al., arXiv preprint, arXiv:1707.08819, 2017.
[6] C. Li et al., in *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

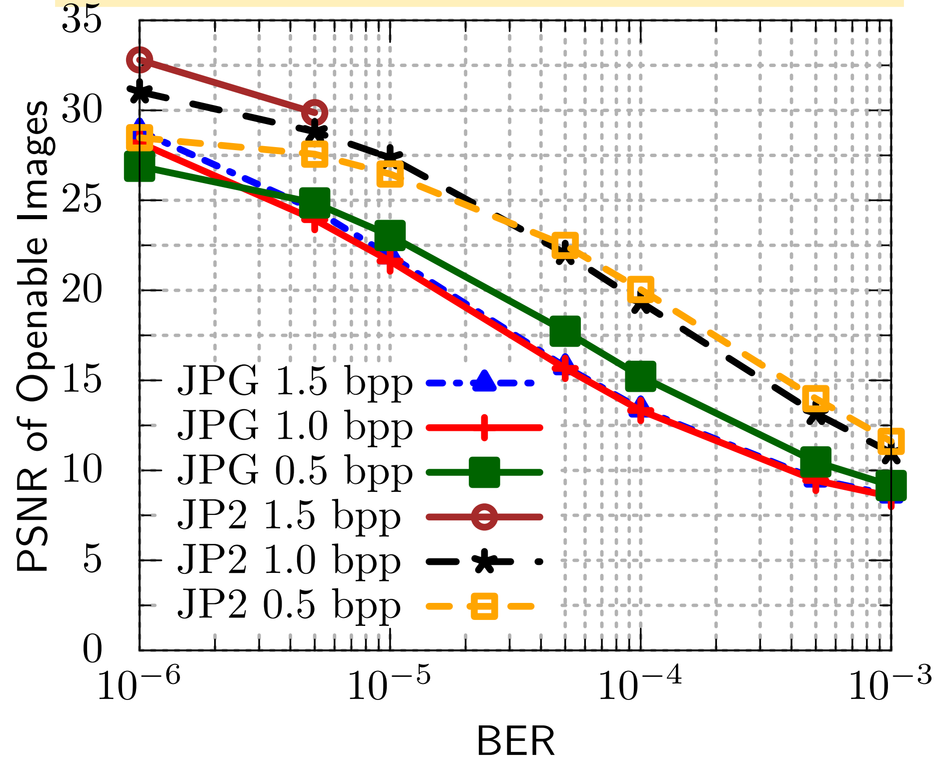
Selection of Baseline Scheme (1)

- Effect of the bit error rate (BER) on the received image quality

Ratio of unopenable images



Received image quality (PSNR)



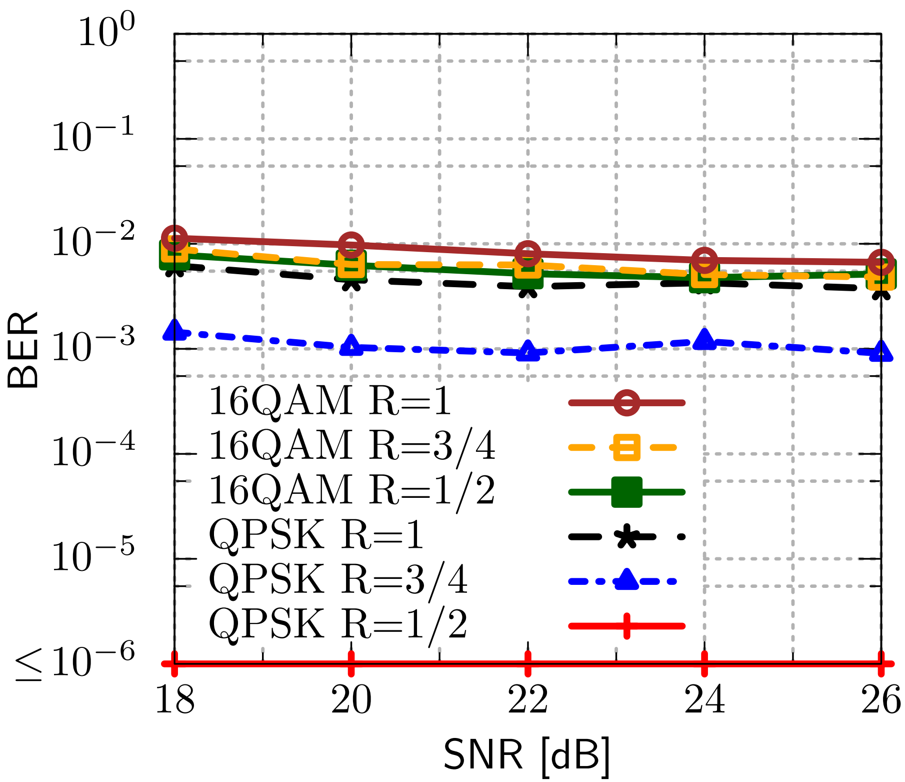
bpp: bit per pixel (compression metric)

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}$$

BER $\leq 10^{-6}$ is required to stably receive JPEG/JPEG2000 images

Selection of Baseline Scheme

- BER curves for QPSK and 16QAM (R: Code rate)



◆ QPSK $(R = 1/2)$:

▪ The only scheme with $\mathrm{BER} \leq 10^{-6}$ (among those in the figure)

➨ The baseline scheme we employ for comparison

In what follows, we suppose

$\mathrm{BER} = 0$ for $\mathrm{SNR} \geq 18\mathrm{dB}$

Performance Evaluation (1)

- Trained the encoder/decoder with 0.5 symbols per pixel (spp)
  - Each 256×192 image is converted to $24,576$ symbols
- The symbol-set size $|S|$ was set to 256
- 70,000 training steps (including 40,000 pre-training steps)

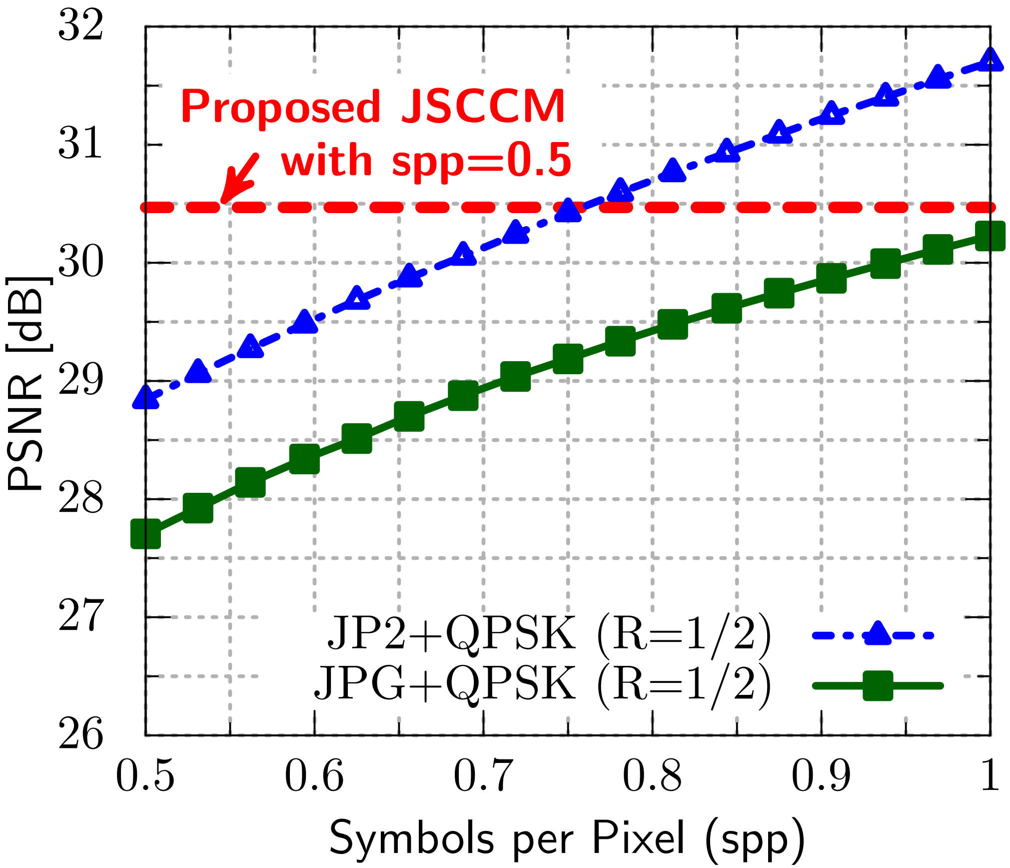
Comparison of the proposed JSCCM with the baseline scheme

- Average PSNR of received images for SNR = 18 dB

Proposed JSCCM	QPSK + JPEG2000	QPSK + JPEG
30.470	28.842	27.706

Performance Evaluation (2)

Comparison of the proposed JSCCM with the baseline scheme



(Figure) Effect of the # of transmitted symbols on the received image quality

- To achieve the same quality as the proposed JSCCM
  - For JPEG2000, we need ~ 50% additional symbols
  - For JPEG, we need ~ 100% additional symbols

➡ **Large improvement in the transmission rate by the proposed JSCCM**

Summary

- Proposed a deep JSCCM scheme for underwater acoustic image transmission, considering
  - Long-delay multipath communication channel
  - Practical transmitter and receiver filters
  - Time-domain adaptive equalizer (TD-AE) to mitigate ISI
- Showed its effectiveness with simulation:
  - **~ 50% speed up** compared with JPEG2000+QPSK
  - **~ 100% speed up** compared with JPEG+QPSK
- Next step: Performance evaluation with real implementation

Example of Received Images

Original

Proposed JSCCM

JPEG2000 + QPSK

JPEG + QPSK

Trained Symbolset