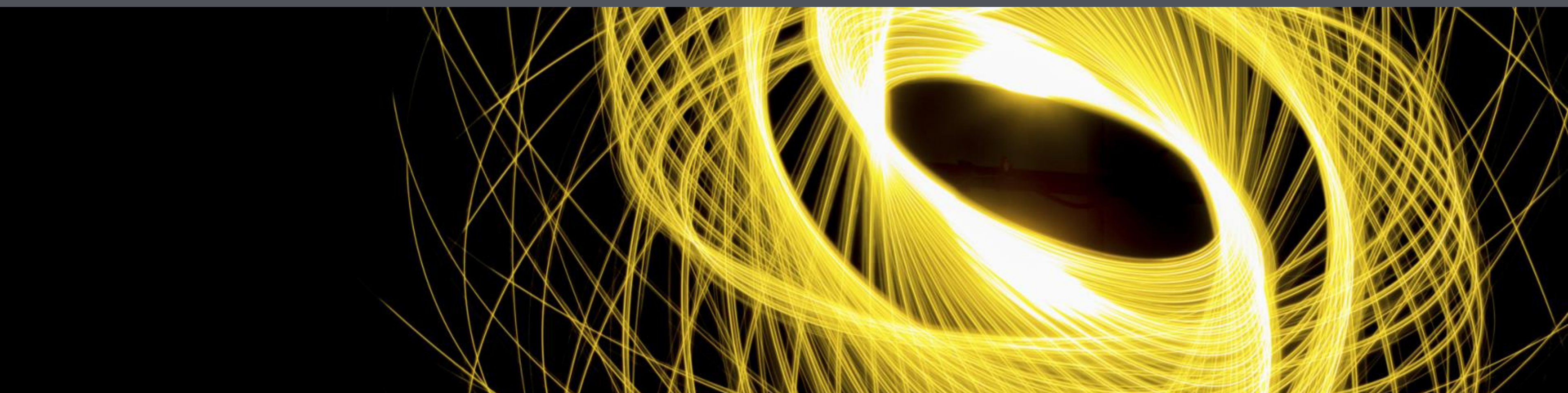
Department of Computer Science

University of Reading

Deep Learning based Communications System with Co-channel Interference



Professor Shuang-Hua Yang, FIET, FInstMC, DSc, PhD
shuang-hua.yang@reading.ac.uk

Outline

- Overview
- Radar-communication co-existence
- Semantic communication for multiple users

Overview - background

- Traditional communication system
  - Bit-based
  - Block-wise optimization
  - Trade-off between system performance and complexity
  - Reaches Shannon capacity limit
- Deep learning based communications system
  - Data-driven
  - End-to-end learning based optimization
  - Global optimization

Overview - background

- Co-channel interference
  - Caused by multiple communications signals transmitted on the same channel or frequency band.
  - Hard to mitigate

Related works

- T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, 2017.
- H. Ye, G. Y. Li, and B. Juang, “Power of deep learning for channel estimation and signal detection in ofdm systems,” IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 114–117, Feb 2018.
- D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, “Machine learning in the air,” IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2184–2199, Oct 2019.
- M. Chen, U. Challita, W. Saad, C. Yin and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3039-3071, Fourthquarter 2019.
- N. Farsad, M. Rao, and A. Goldsmith, “Deep learning for joint source-channel coding of text,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 2326–2330.
- Z. Qin, H. Ye, G. Y. Li, and B. F. Juang, “Deep learning in physical layer communications,” IEEE Wireless Communications, vol. 26, no. 2, pp. 93–99, April 2019.
- H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” IEEE Transactions on Signal Processing, vol. 69, pp. 2663–2675, 2021

Overview - motivation

- Challenges in learning based communications system
  - Co-channel interference (avoid, mitigate or tolerate?)
    - Operate independently
    - Operate over the same frequency and time

  - Learning based model
    - Generalization ability to different channel conditions
    - Limited training datasets
    - Model size

Overview - focus

- In this presentation:
  - Radar-communication co-existence [1]
    - Unsupervised Principle Component Analysis (PCA) assisted deep learning based signal detector
  - Semantic communication [2]
    - Knowledge distillation (KD) for generalizability enhancement

[1] C. Liu, Y. Chen and S. -H. Yang, "Deep Learning Based Detection for Communications Systems With Radar Interference," in IEEE Transactions on Vehicular Technology, June 2022
[2] C. Liu, Y. Chen and S. -H. Yang, "Knowledge distillation based semantic communications for multiple users," in IEEE Transactions on Wireless Communications, under review

Outline

- Overview
- **Radar-communication co-existence**
- Semantic communication with co-channel interference
- Cooperative perception in autonomous driving

C. Liu, Y. Chen and S. -H. Yang, "Deep Learning Based Detection for Communications Systems With Radar Interference," in *IEEE Transactions on Vehicular Technology*, June 2022

Radar-communication co-existence

- Carrier frequency has been moving towards radar band because of increasing transmission rate
- Radar has been utilized in civilian application, such as traffic control, vehicle cruise

- Radar-communication co-existence is proposed to improve spectrum efficiency.

Radar-communication co-existence

- Generally , there are two types of co-existence
  - **Operate independently**
    - Resource allocation and co-design of both systems are required to cooperatively manage and share the spectrum
    - No mutual interference
  - **Operate over the same time and frequency**
    - Cause mutual interference due to spectral overlap
    - Redesign one system to better withstand the other
    - Interference mitigation required

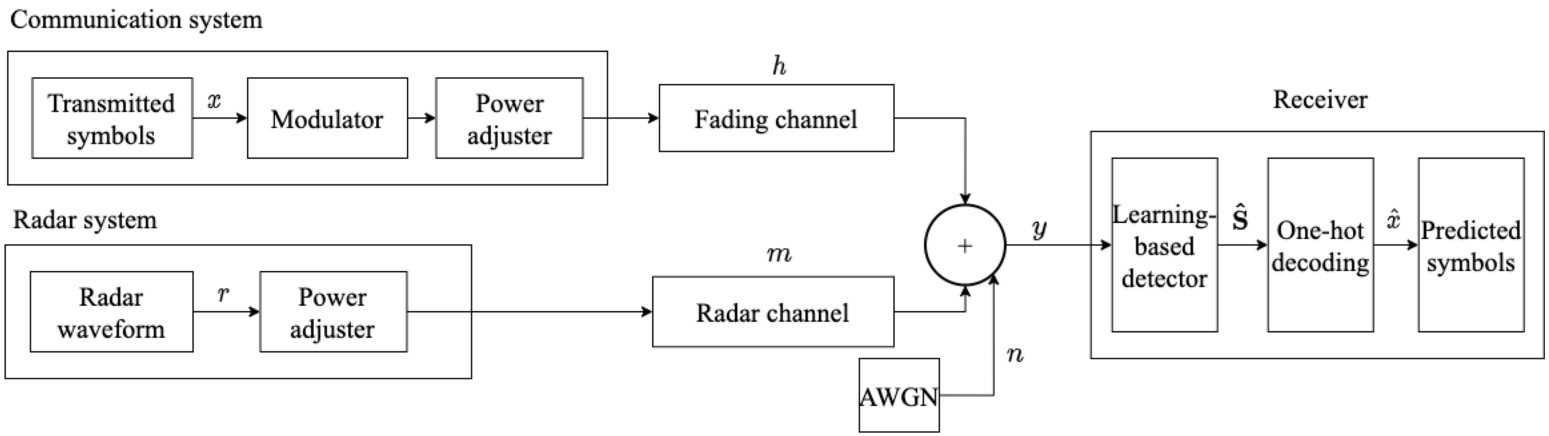
Radar-communication co-existence

- Learning based signal detection
  - Leverage neural networks' pattern recognition ability
  - Directly recover signals from observed symbols
  - No need to mitigate interference

  - However, Offline training is required

System model

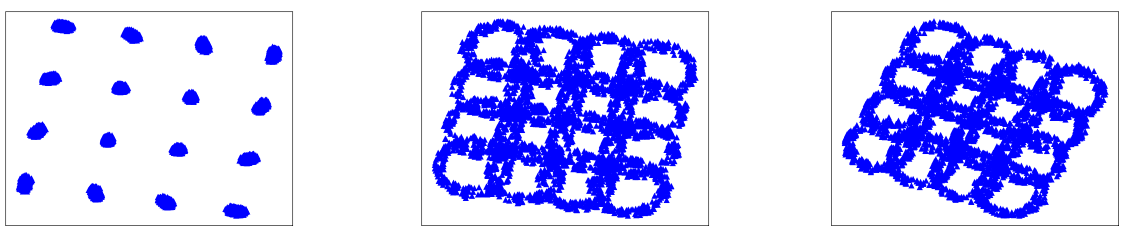
- Radar-communication coexistence
  - Channel model : $y = hx + mr + n$



Our goal is to recover x fromreceived y in the presence of radar interference r

Learning based signal detector

- Preprocessed by Principal Component Analysis (PCA)
  - Originally for dimensionality reduction
  - Learn the best rotation angle based on the density of points, and transform to new coordinate system, to make symbols less intersect.



(a) (b) (c)

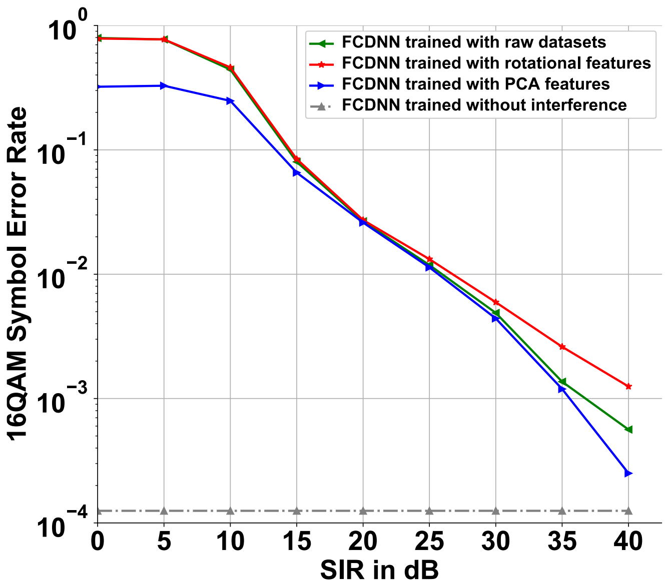
Constellations of 16QAM signals in the presence of radar interference
(a) SIR = 30 dB (b) SIR = 10 dB (c) SIR = 10 dB with preprocessing by PCA.

Learning based signal detector

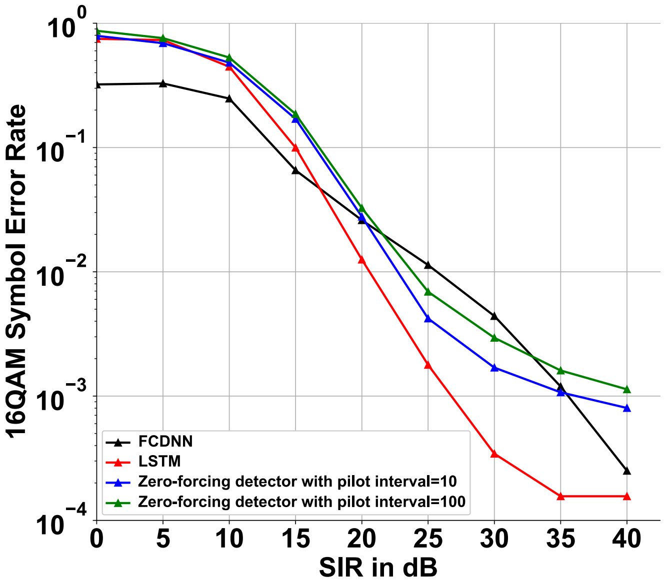
- Learning based detector:
  - **Fully connected deep neural networks (FCDNN)**
    - Symbol-by-symbol detector
    - Input : Symbol and PCA features
    - Output: a predicted symbol
  - **Long short-term memory (LSTM)**
    - Sequence detector
    - Input : a sequence of symbols and PCA features
    - Output: a sequence of predicted symbols

Simulation results

- Performance in the presence of radar interference
  - **Better performance for FCDNN** with PCA features.
  - Generally, **learning-based detector** performs **better**.



Symbol error rate of 16QAM signals with detectors having different features when the SNR is 40 dB.



Symbol error rate of 16QAM signals in the presence of LFM interference when the SNR is 40 dB.

Outline

- Overview
- Radar-communication co-existence
- **Semantic communication with co-channel interference**
- Cooperative perception in autonomous driving

Semantic communication

Based on Shannon and Weaver, communication can be categorized into:

- Level 1: Transmission of symbols (bits)
- **Level 2: Semantic information exchange (semantics)**
- Level 3: Effectiveness of information exchange (tasks/goals)

Semantic communication focus on the precision of the recovered information, instead of aiming for the accuracy at bit level

C. E. Shannon and W. Weaver, “The mathematical theory of communication,” The University of Illinois Press, 1949.

Related works

- H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," IEEE Transactions on Signal Processing, 2021
  - Proposed **DeepSC**, a transformer based semantic communication model

- H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," IEEE Journal on Selected Areas in Communications, 2021
  - Proposed **L-DeepSC**, a light-weight DeepSC with small size

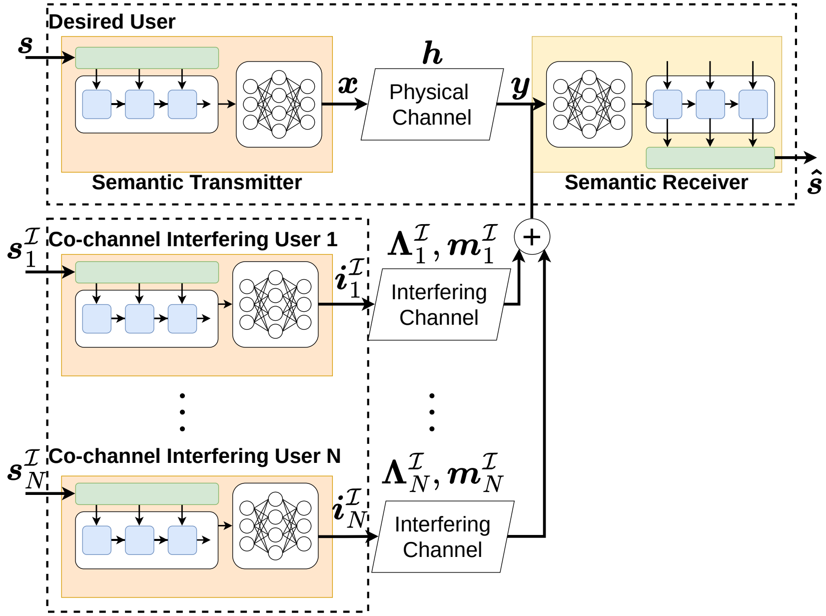
- G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in NIPS Deep Learning and Representation Learning Workshop, 2015.
  - First proposed and adopted knowledge distillation in neural network

System model

Based on DeepSC, we proposed a semantic communication for multiple users,

- Semantic encoder-decoder
- Channel encoder-decoder
- Co-channel semantic interference
- Loss function:

$$\mathcal{L}_{hard} = \mathcal{L}_{CE}(\boldsymbol{s}, \hat{\boldsymbol{s}}) = -\frac{1}{n}\sum_{m=1}^{n} P(\boldsymbol{s}[m]) \log P(\hat{\boldsymbol{s}}[m])$$



$$\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \sum_{k=1}^{N} \boldsymbol{\Lambda}_k^{\mathcal{I}} * \boldsymbol{m}_k^{\mathcal{I}} * \boldsymbol{i}_k^{\mathcal{I}} + \boldsymbol{n},$$

$$\boldsymbol{\Lambda}_k^{\mathcal{I}} = [\lambda_{k,1}, \lambda_{k,2}, \cdots, \lambda_{k,n}],$$

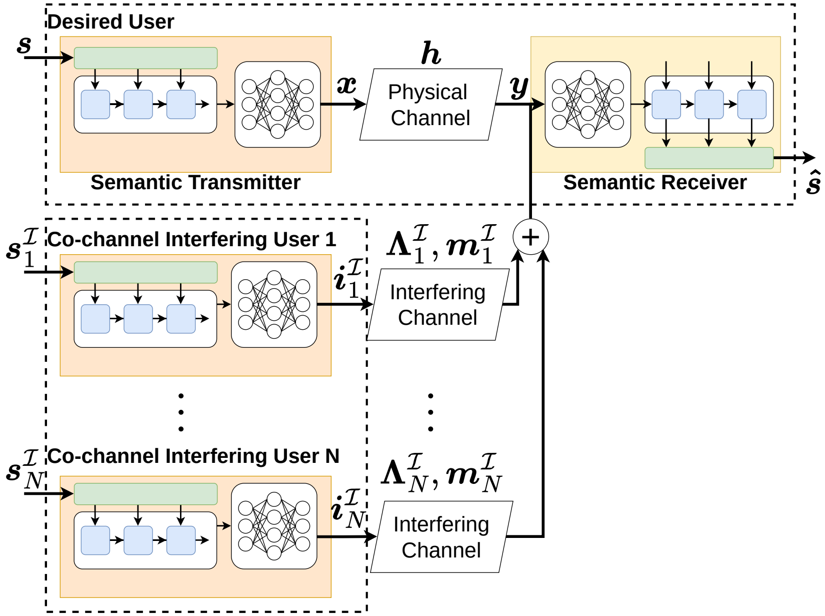
$$\lambda_{k,j} = \begin{cases} 1 & \text{if the interference symbol occurs} \\ 0 & \text{otherwise} \end{cases},$$

Challenge

- How well can the model **generalize** on **unseen interference**?
  - Easy to **overfit** on the training data

- How well can the model perform by training with **limited** dataset?
  - Difficult to obtain well-labelled dataset with **all patterns included**.

- How **light** can the model be with negligible performance loss?
  - **Model complexity**

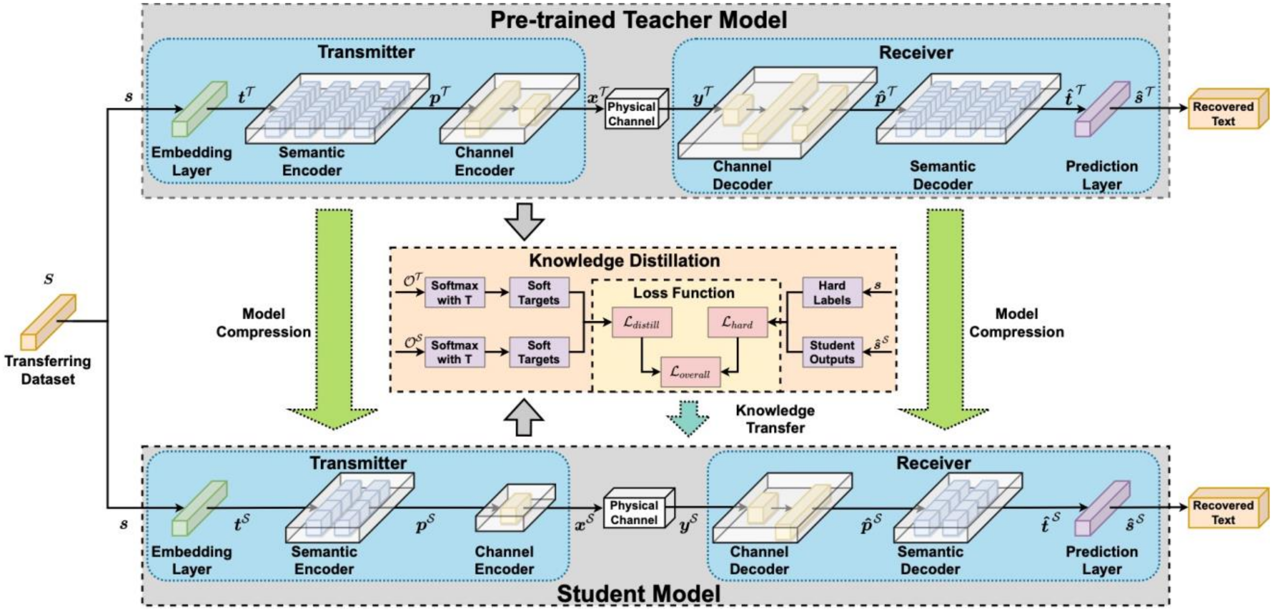


Motivation

- Why Knowledge distillation ?
  - **Learning from Teacher** VS **Learning from labelled data**
    - Teacher's soft information contain more generalized features
    - Labelled data only contain hard label information
    - Expensive to obtain well labelled data
  - **Model compression** without losing **generalizability**
  - Potentially **more control** on the optimization of **different blocks**
  - **Better data privacy**

Proposed method

- Knowledge distillation for semantic communications with co-channel interference
  - Teacher :Pretrained, large model, over-parameterized
  - Student : Light model, guided by Teacher



Proposed method

- Knowledge distillation for semantic communication
  - Knowledge : Soft information computed by $Q(z_i;T) = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$
  - Distillation loss:
    - KL divergence between Student and Teacher

$$\mathcal{L}_{total\_distill} = \sum_{(\mathcal{O}^{\mathcal{S}},\mathcal{O}^{\mathcal{T}})\in\mathbb{O}} \mathcal{L}_{distill}(\mathcal{O}^{\mathcal{S}},\mathcal{O}^{\mathcal{T}})$$

$$= \sum_{(\mathcal{O}^{\mathcal{S}},\mathcal{O}^{\mathcal{T}})\in\mathbb{O}} \mathbb{E}\left\{\eta_{\mathcal{O}^{\mathcal{S}},\mathcal{O}^{\mathcal{T}}} T^2 \mathcal{D}_{KL}[Q(\mathcal{O}^{\mathcal{S}};T)||Q(\mathcal{O}^{\mathcal{T}};T)]\right\},$$

$$\mathbb{O} \subseteq \left\{(\boldsymbol{p}^{\mathcal{S}},\boldsymbol{p}^{\mathcal{T}}),(\boldsymbol{x}^{\mathcal{S}},\boldsymbol{x}^{\mathcal{T}}),(\hat{\boldsymbol{p}}^{\mathcal{S}},\hat{\boldsymbol{p}}^{\mathcal{T}}),(\hat{\boldsymbol{t}}^{\mathcal{S}},\hat{\boldsymbol{t}}^{\mathcal{T}}),(\hat{\boldsymbol{s}}^{\mathcal{S}},\hat{\boldsymbol{s}}^{\mathcal{T}})\right\},$$

Proposed method

- Knowledge distillation for semantic communication
  - Overall loss:
    - Combining hard-label cross-entropy loss and distillation loss

$$\mathcal{L}_{overall} = \Big(1 - \sum_{(O^{\mathcal{T}}, O^{\mathcal{S}}) \in \mathbb{O}} \eta_{O^{\mathcal{T}}, O^{\mathcal{S}}}\Big) \mathcal{L}_{CE}(\boldsymbol{s}^{\mathcal{S}}, \hat{\boldsymbol{s}}^{\mathcal{S}})$$

$$+ \sum_{(O^{\mathcal{T}}, O^{\mathcal{S}}) \in \mathbb{O}} \mathbb{E}\Big\{\eta_{O^{\mathcal{T}}, O^{\mathcal{S}}} T^2 \mathcal{D}_{KL}(Q(O^{\mathcal{S}}; T) || Q(O^{\mathcal{T}}; T))\Big\}.$$

Proposed method

- Knowledge distillation for semantic communications
  - Model compression
    - Reduce the redundant layers
    - Post-dynamic quantization for model weights

Simulation

- Learning based baselines
  - DeepSC [1]
  - L-DeepSC [2] structure as Teacher, Student 1 and Baseline 1
  - Baseline 1, 2, 3
    - Same structure as Student 1, 2, 3 but without knowledge distillation
- Conventional communication systems
  - Huffman and LDPC
  - 5-Bit and LDPC

TABLE I: The setting of the distilled knowledge for the student models

	Student 1	Student 2	Student 3	Student 4
Semantic encoder	-	-	✓	✓
Channel encoder	✓	✓	✓	✓
Channel decoder	-	-	✓	✓
Semantic decoder	✓	✓	✓	✓
Prediction layer	✓	✓	✓	✓

TABLE II: The setting of the model compression for the student models

	Student 1	Student 2	Student 3	Student 4
Semantic encoder	-	✓	✓	✓
Channel encoder	-	-	✓	✓
Channel decoder	-	-	✓	✓
Semantic decoder	-	✓	✓	✓
Prediction layer	-	-	-	-
Quantization	-	-	-	✓

[1] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," IEEE Transactions on Signal Processing, 2021
[2] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," IEEE Journal on Selected Areas in Communications, 2021

Simulation

- Teacher
  - Pretrained with extensive dataset
- Student
  - Limited training dataset with limited regime of SNR, no interference samples
  - With knowledge distillation
- Baselines
  - Same datasets with Students
  - Without knowledge distillation

TABLE I: The setting of the distilled knowledge for the student models

	Student 1	Student 2	Student 3	Student 4
Semantic encoder	-	-	✓	✓
Channel encoder	✓	✓	✓	✓
Channel decoder	-	-	✓	✓
Semantic decoder	✓	✓	✓	✓
Prediction layer	✓	✓	✓	✓

TABLE II: The setting of the model compression for the student models

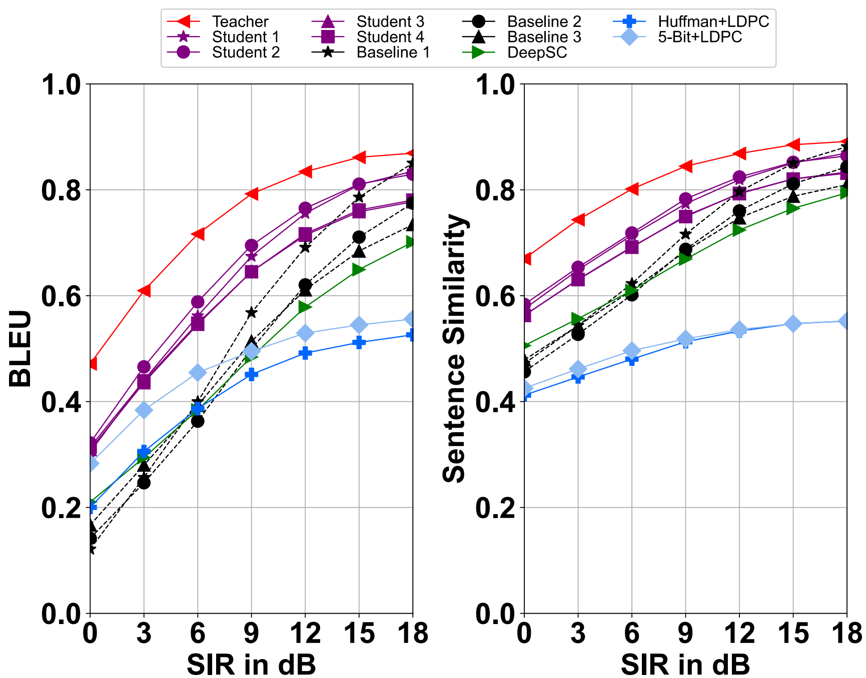
	Student 1	Student 2	Student 3	Student 4
Semantic encoder	-	✓	✓	✓
Channel encoder	-	-	✓	✓
Channel decoder	-	-	✓	✓
Semantic decoder	-	✓	✓	✓
Prediction layer	-	-	-	-
Quantization	-	-	-	✓

Simulation

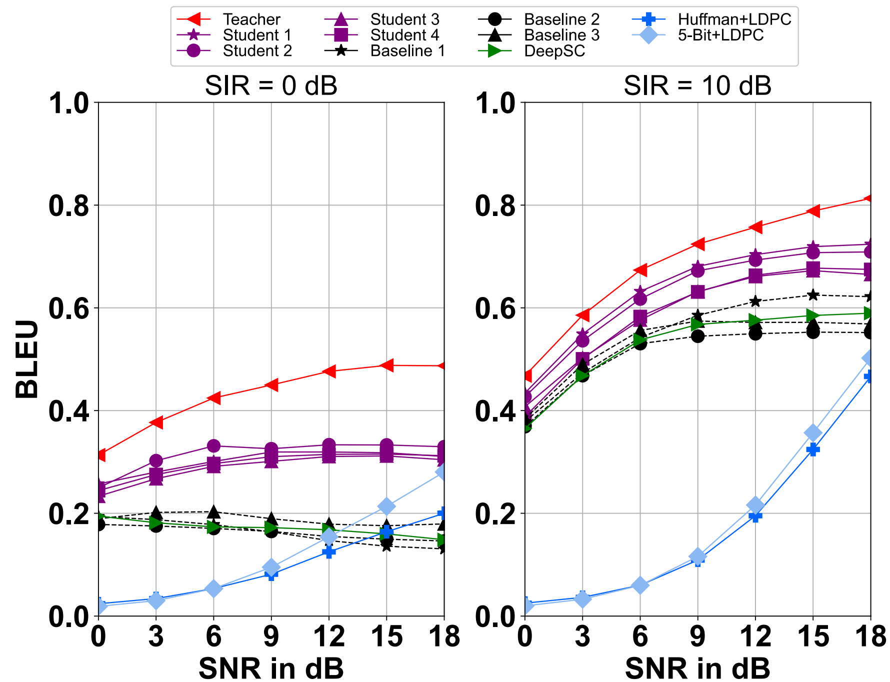
- Performance metrics
  - BLEU [1]
    - Compare the sentence differences
    - Difficult to distinguish synonyms or polysemy
  - Sentence similarity [2]
    - Sentences with similar meaning tend to have closer vector distance
    - Use pretrained BERT to map the sentences into semantic vector space and compare their semantic vectors.

[1] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Jul. 2002
[2] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” IEEE Transactions on Signal Processing, 2021

Simulation results



BLEU score and sentence similarity of the semantic communications with one multi-user interference when the SNR is 18 dB



BLEU score of the semantic communications with one multi-user interference

Simulation results

- Complexity analysis
  - Knowledge distillation leads to an increase of training time
  - Model compression reduces the model size and inference time

	Parameters	Size (MB)	Training time (ms/batch)	Inference time (ms/sentence)
Teacher	2022672	12.46	108.86	24.10
Student 1	2022672	12.46	199.75	22.90
Student 2	1096976	6.98	169.87	14.87
Student 3	946704	6.06	166.86	14.50
Student 4	5376	0.05	166.86	14.18
DeepSC	1462928	9.18	95.76	19.79
LDPC	-	-	-	42.74

Conclusion

- Distilled models perform better than the non-distilled baselines and the conventional communications system

- Knowledge distillation can reduce performance loss while compressing the model

- Post-training dynamic quantization has a very limited effect on the system performance

Thank you