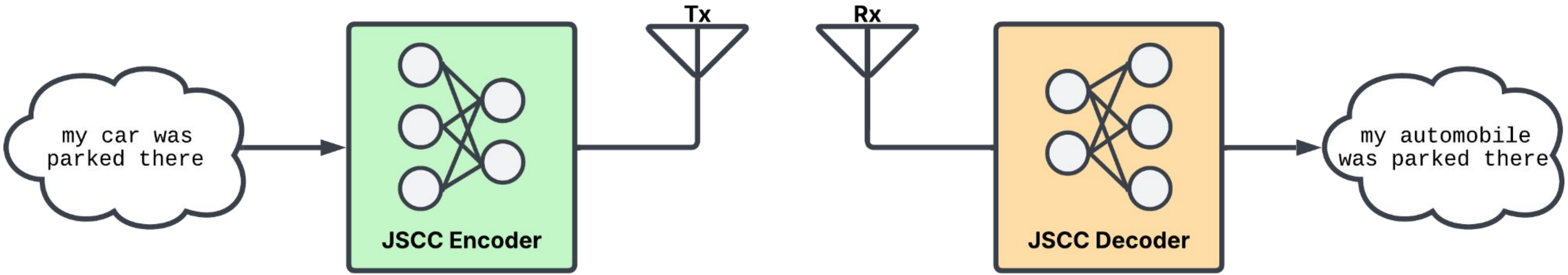
JSCC for Semantic Text Transmission: Comm with Built-in Translation

Jason Ning, Joshua Ning
zyning@umich.edu joshning@umich.edu

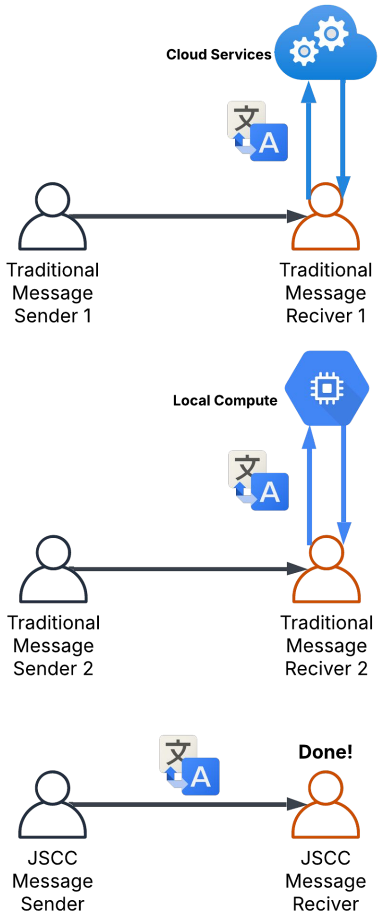
Background

- JSCC is a unified method to combine compression and channel coding
- Deep JSCCs for text capture semantics in sentences
- Current state-of-the-art model [1] uses Transformers
    - Performs well in low SNR environments
    - Suffers from semantic distortion



Motivation

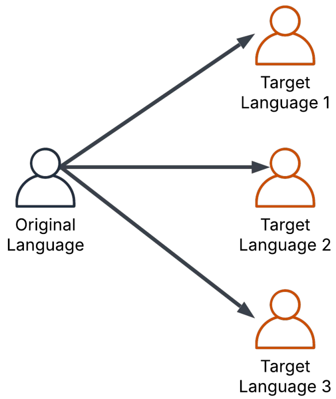
- Semantic text communication is unsatisfactory
    - Most users want to receive accurate text data
- Semantic distortion is tolerable in translation applications
- JSCC can directly decode message to another language
    - Faster than conventional methods (cloud & local)
    - 1 encoder → many language decoder simultaneously
    - A novel application (to the best of our knowledge)



Project Goal

Minimal Viable Product

- Demonstrate translation ability in simulation ✅
- Simple AWGN Channel ✅
- Demonstrate single encoder working with decoders with different target languages ✅



Reach Goals:

- Implementation on the USRP in conjunction with small GPU❌

- Mainly due to lack of time, should be easily deployable

System Architecture (DeepSC)

- Encoder: Text -> embedding -> transformer -> Dense layer.
- Decoder: Dense layer -> transformer -> embedding -> Text.
- Additional networks for mutual information estimation between X and Y to help channel encoder converge (implemented but not used)
- Original paper used a single language for the text ie, english -> english

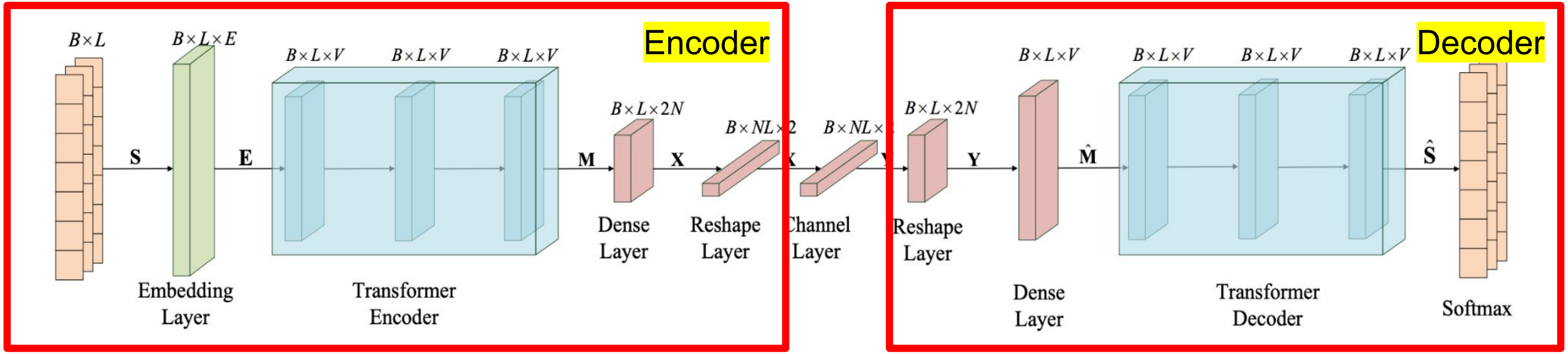
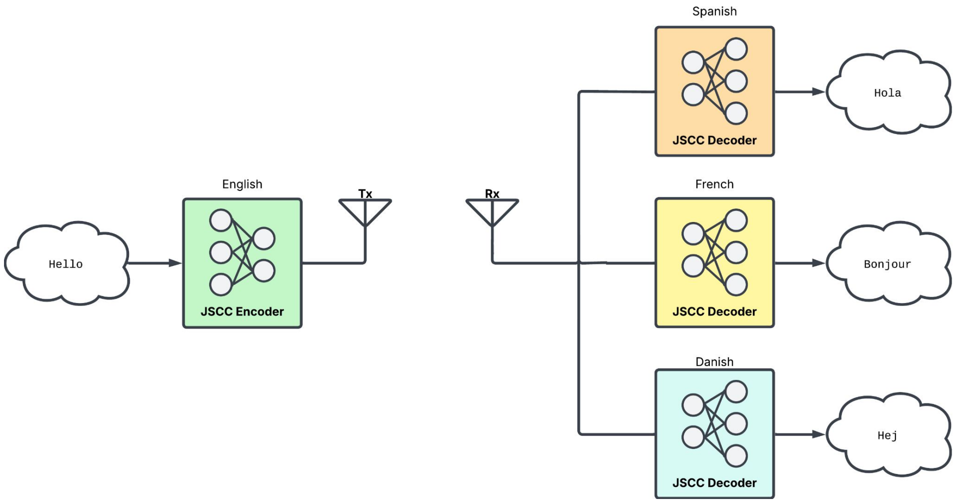


Fig. 2. The proposed neural network structure for the semantic communication system.

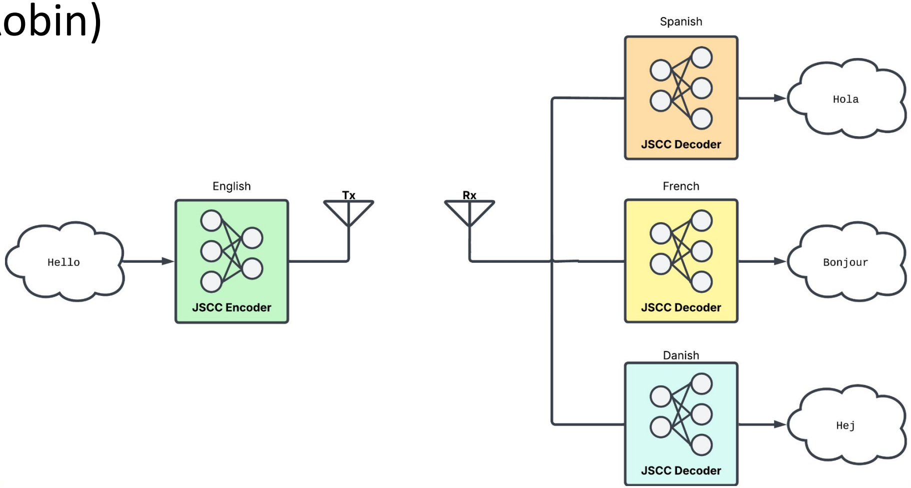
Our Architecture

- One encoder to many decoder, each corresponding to a different language
- During inference user can choose which decoder to use
- 1 to many languages should help encoder better capture semantic information



Training Method

- Construct datasets for each language pairs
  - English <-> Spanish, English <-> French, …
- At the beginning of each epoch:
  - Select a decoder and it's dataset (Round Robin)
  - Train the encoder and selected decoder
  - Send decoder back to CPU
  - Evaluate all models after 1 cycle of RR

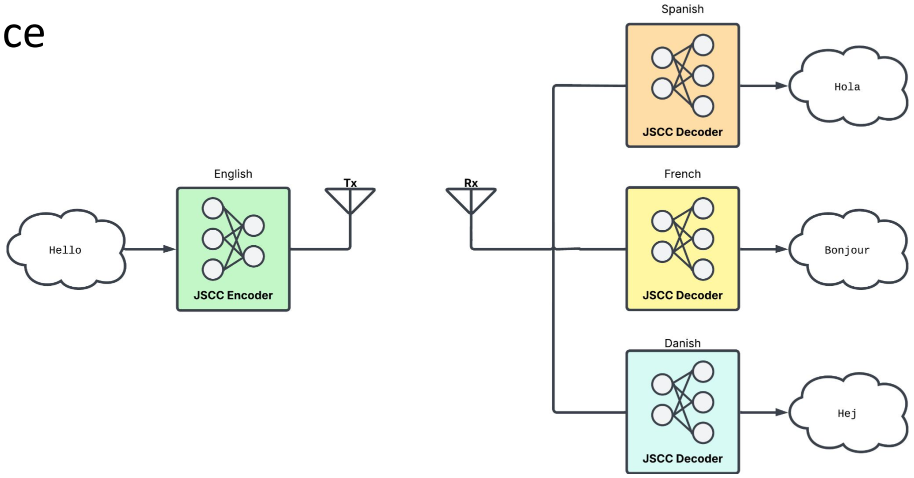


Training Method Rationalities

- Initially, we want to train all the decoders at once
  - Pro: Possibly better and faster convergence
  - Con:

    Require large amounts of compute

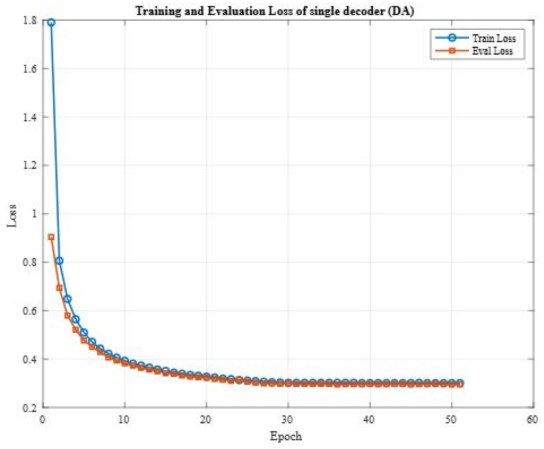
    Require large multilingual dataset

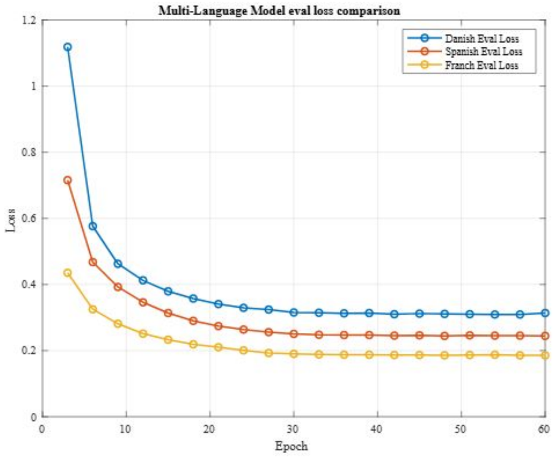


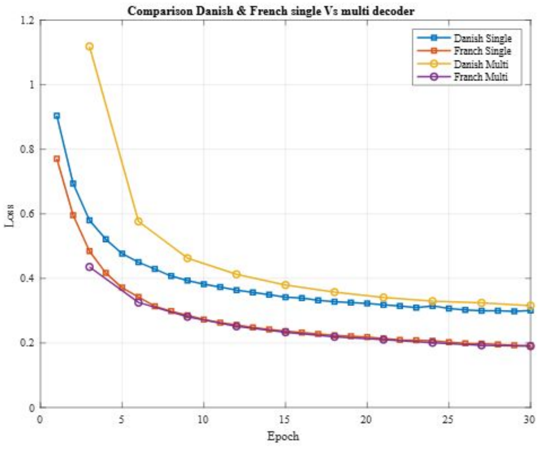
- Why round robin for decoder per epoch?
  - Tried round robin per batch -> too slow
  - The language trained closest to the evaluation will have advantage

Results

- Our implementation of DeepSC used for translation(1 encoder 1 decoder)
- Our Training Method (1 encoder 3 decoder)
- Our method achieve lower cross entropy loss while obtaining 3 decoders







Results: Sample Translations at SNR 0.1dB

Sentence 1: 100% success (majority)

src lang = <START> in kosovo , it is about to happen all over again . <END>

trg lang = <START> c ' est ce qui risque de se passer à nouveau avec le kosovo . <END>

trg lang gt = <START> c ' est ce qui risque de se passer à nouveau avec le kosovo . <END>

Sentence 2:

src lang = <START> initially , the lowest performing producers will be forced to stop producing and the level of european production will therefore fall . <END>

trg lang = <START> dans un premier temps , les producteurs les moins performants seront amenés à disparaître et nous erika ainsi le niveau de production européen . <END>

trg lang gt = <START> dans un premier temps , les producteurs les moins performants seront amenés à disparaître et nous diminuerons ainsi le niveau de production européen . <END>

Note: one of the most important word is replaced by a person’s name (Erika).

Results: Sample Translations at SNR 0.1dB

Sentence 3:

src lang = <START> put the eu on a diet , and give greater freedom to democracy in our countries !

trg lang = <START> faisons suivre une lomé d ' résout à l ' ue et accordons plus de liberté aux démocraties de nos pays ! <END>

trg lang gt = <START> faisons suivre une cure d ' amaigrissement à l ' ue et accordons plus de liberté aux démocraties de nos pays ! <END>

Note : The sentence is grammatically broken and does not make sense. But important info are preserved.

Conclusion

- Text based JSCC still has limitations. Semantics distortion is not the biggest source of error (at least on our dataset)
  - Car -> automobile distortion is rare
  - Word -> gibberish is more common
- Main contribution are:
  - First to demonstrate Deep JSCC works with translation purposes
  - Introduce a new training method that can speed up training process and increase modularity of text JSCC
  - Faster convergence even with multi language training (3x training speed

Challenges

- Compute limitations: train & eval on RTX 4070 12 GB
- Original paper result was somewhat cherry picked
- Can’t obtain large multilingual dataset
- Long training time: 5-10 minutes per epoch.
- Multi-decoder architecture require unloading from GPU (tricky to figure out)

Sources

[1] https://arxiv.org/pdf/2006.10685

[2] https://ieeexplore.ieee.org/document/9714510

Questions and Discussion