

KIT's Multilingual Neural Machine Translation systems for IWSLT 2017

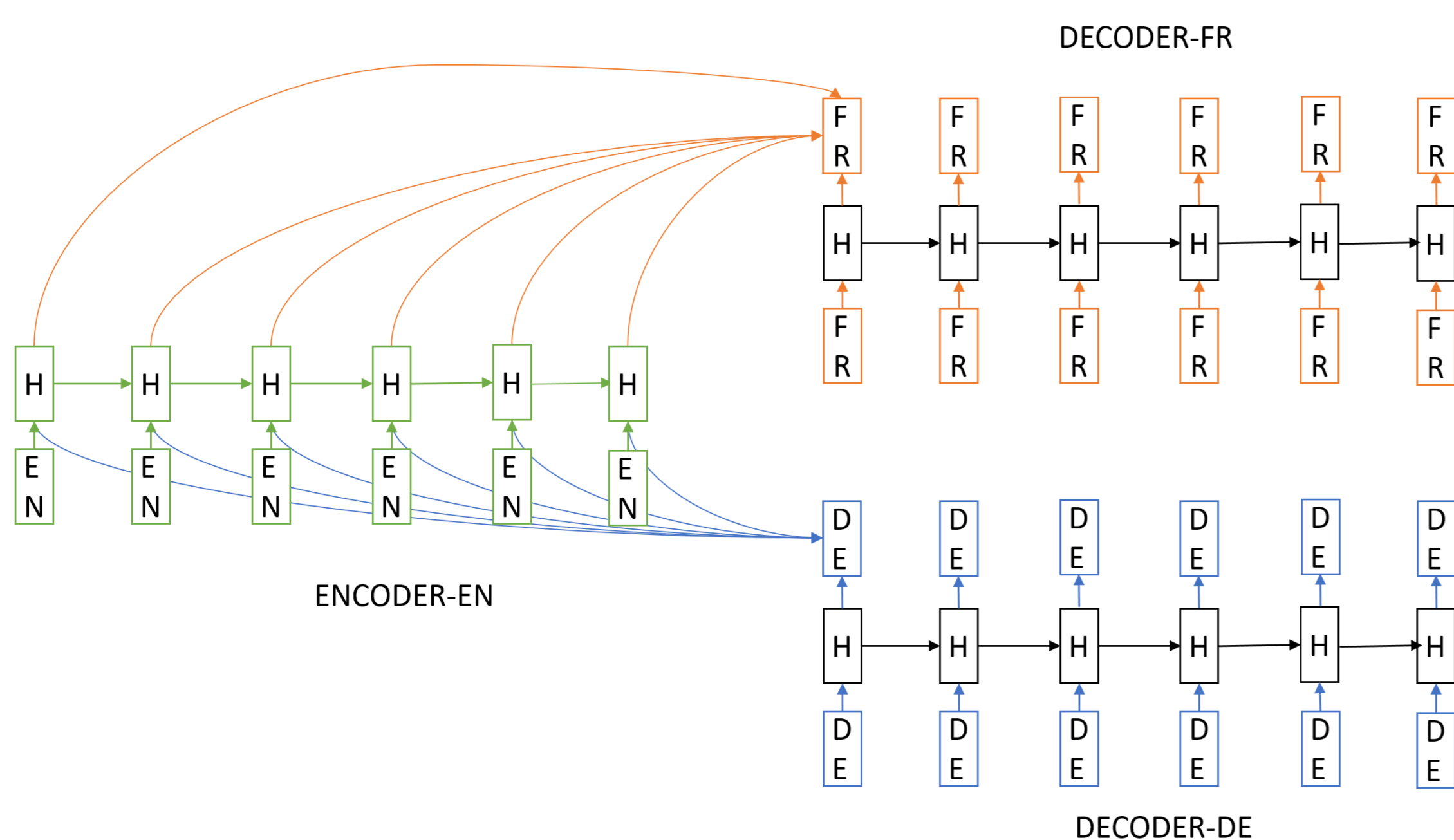
Ngoc-Quan Pham, Matthias Sperber, Elizabeth Salesky, Thanh-Le Ha, Jan Niehues, Alexander Waibel

Overview

- Neural machine translation applied for multi-lingual setting
- One model - Multiple directions
- Architecture comparison
- Adaptation improves direction-wise performance
- Can the model utilise large data ?
- Can we apply the model in a Zero-Shot setting ?

Multi-lingual architectures

- Embedding: One space per language - reusable across encoders/decoders
- Encoders: Each language is encoded using one bi-directional recurrent network
- Decoders: One recurrent network for each language
- Output layers (Projection & Softmax): Distinctive per language
- Attention mechanism: one layer per language-pair
- Dynamic computational graph implementation: Only use network components involved in a particular language pair

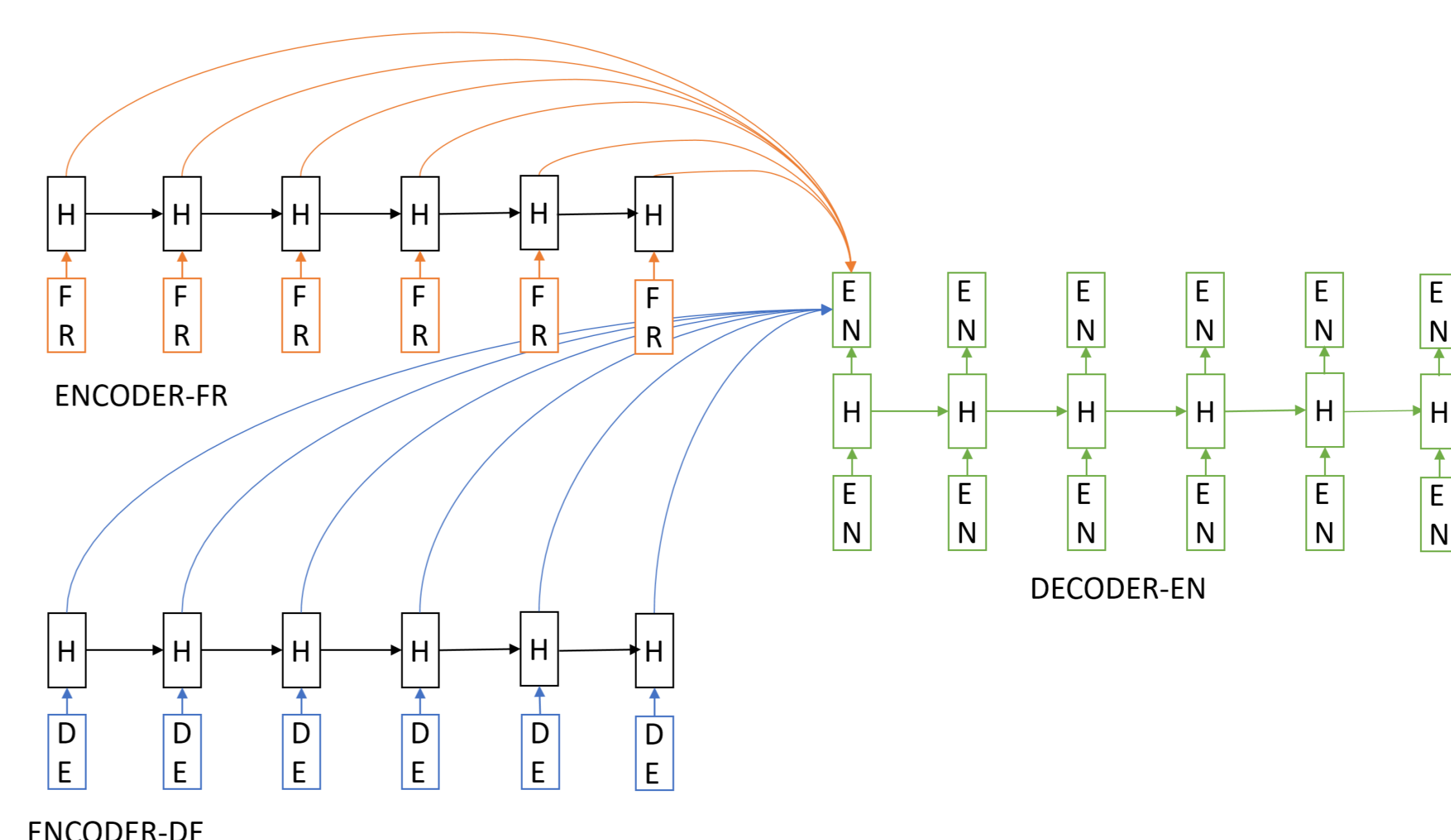


Sharing-Strategy

- Share Everything: one single set of model parameters learn from multiple languages
- Share RNN: One encoder and one decoder for all languages
- Share Attention: One attention layer for all language pairs
- Separate All: All components are language-specific

Adaptation Strategy

- Training on large data - Adapt on TED talks
- Fine-tuning the models on each direction
- Averaging the parameters of language-specific adapted models



Results

Small Task

System	tokenized BLEU	case-sensitive BLEU
Separate-All	24.7	22.6
+ Lang-adapted	25.8	24
Share-RNN	26.0	24.2
+ Lang-adapted	26.3	24.5
Share-All	25.2	23.5
+ Lang-adapted	26.2	24.2
Language-coded	25.6	23.8
Share-All Average	25.7	23.8
Ensemble	27.4	25.6

Speech translation Task

System	tst2013 EN-DE	tst2013 DE-EN
baseline	17.9	15.7
noise	18.4	16.0

Large Task

System	tokenized BLEU	case-sensitive BLEU
Share-All	26.9	25.1
Share-RNN	26.9	25.1
Separate-All	25.1	23.4
Ensemble	27.8	26.0

Zero Resource

System	DE→NL		NL→DE	
	dev2010	tst2010	dev2010	tst2010
Zero	15.87	19.46	14.03	19.59
Zero Filtered Dict	15.79	19.48	13.96	19.59
Zero Lang Feature	16.65	19.68	14.50	20.67
System	IT→RO		RO→IT	
	dev2010	tst2010	dev2010	tst2010
Zero	11.61	15.44	16.18	17.11
Zero Filtered Dict	11.52	15.45	16.21	17.20
Zero Lang Feature	12.70	16.22	17.26	17.79