

## 1 Sequence to Sequence Learning

- Encoder-decoder architecture: widely applicable to many tasks
  - Machine translation.
  - Image captioning.
  - Constituent parsing.
  - Unsupervised learning
- These tasks are almost always learned in *isolation*!

Can we benefit from multi-task seq2seq learning?

## 3 Experiments

- Translation:** (Luong et al., 2015)
  - WMT English  $\rightleftharpoons$  German: 4.5M examples.
- Parsing:** (Vinyals et al., 2015a)
  - Penn Tree Bank (PTB): 40K examples.
  - High Confidence (HC): 11M examples.
- Caption:** (Vinyals et al., 2015b)
  - 600K examples.
- Unsupervised:** auto-encoders & skip-thought
  - 12.1M English and 13.8M German examples.
- Setup:** (Sutskever et al., 2014), *attention-free*
  - 4-layer deep LSTMs: 1000-dim cells/embeddings.

## 5 Parsing Results

Big (HC parsing) + Big (translation)

Systems	Parsing F <sub>1</sub>	attention
LSTM+A (Vinyals et al., 2015a)	92.5	3 layers 256 cells 512 embs
LSTM+A+Ensemble 6 (Vinyals et al., 2015a)	<b>92.8</b>	
<i>Our systems</i>		
HC Parsing	92.2	no attention 4 layers 1000 cells 1000 embs
HC parsing + translation (0.1x)	92.0	
HC parsing + translation (0.05x)	<b>92.4</b>	
HC parsing + translation (0.01x)	92.2	
Ensemble of 6 models	<b>93.0</b>	

Recent LM results (Jozefowicz et al., 2016): big LSTMs win!

**New SOTA**

## 2 Multi-task Learning Approaches

- One-to-many: shared encoder**
- Many-to-one: shared decoder**
- Many-to-many: shared encoder & decoder**

- Alternative training:**
  - One task per mini-batch.
  - Different sampling frequencies

Translation improvement with little parsing / caption data.

- State-of-the-art constituent parsing.

## 4 Translation Results

Big (translation) + Small (PTB parsing)

English-German Translation (BLEU)

Mixing ratio

Big (translation) + Medium (caption)

German-English Translation (BLEU)

## 6 Translation + Unsupervised Learning

German-English Systems	Ppl	BLEU
(Luong et al., 2015)	14.3	16.9
Translation	12.5	17.8
<i>Autoencoders</i>		
Translation + autoencoders (1x)	13.9	16.6
Translation + autoencoders (0.05x)	12.0	<b>18.3</b>
<i>Skip-thought Vectors</i>		
Translation + skip-thought (1x)	<b>10.8</b>	17.3
Translation + skip-thought (0.01x)	12.2	17.8

- Autoencoders:** improve translation (+0.5 BLEU).
- Skip-thought:** best perplexity.
  - But **failed** to improve translation!
  - Analysis: tend to produce <unk>.

## 7 Conclusion

- Multi-task learning help:
  - Up to +1.5 BLEU for translation.
  - SOTA constituent parsing
- Future work:** better unsupervised learning metric.