**Sheffield Submissions**  
**For WMT18 Multimodal Translation Task**

Chiraag Lala, Pranava Madhyastha, Carolina Scarton and Lucia Specia  
{clala1, p.madhyastha, c.scarton, l.specia}@sheffield.ac.uk

---

**Task 1**  
Single Source Multimodal Machine Translation  
EN + 📷 → DE or FR or CS

**Re-ranking with WSD**  
A baseline NMT system (SHEF_Base) generates n-best translation hypotheses. These are re-ranked using novel cross-lingual Word Sense Disambiguation (WSD) models.

**SHEF_Base model**  
An ensemble of different runs of a standard attentive NMT model with five different seeds.

10-best translation hypotheses with likelihood scores are generated. It’s 1-best (no re-ranking) forms the SHEF_Base

**Cross-lingual WSD models**

1. **Most Frequent Sense (SHEF_MFS)**  
   \( \text{Freq}_{\text{wooden}}(\text{bois}) = 79 \), \( \text{Freq}_{\text{forest}}(\text{forêt}) = 10 \)  
   Most Frequent Translation of woods is bois

2. **Lexical Translation (SHEF_LT)**  
   people walking down a trail in the woods

3. **Multimodal LT (SHEF_MLT)**  
   These models are trained on the Multimodal Lexical Translation Dataset (https://github.com/sheffieldmlp/mlt) derived from the Multi30K corpus. Caveat: EN-CS is noisy.

**Concatenation and Consensus**  
First, we train standard attentive NMT models for three language directions into Czech and generate 10-best lists in Czech from each model.  
**SHEF_MLT**: concatenate the 10-best lists and then re-rank using MLT model.  
**SHEF_Con**: consensus of the 10-best lists, i.e. select the translation hypothesis that appears in the intersection of the three lists.

**Augmentation and Classifiers**  
First, we add more training data by translating German, French, Czech training instances into English and then train EN-CS NMT model.  
The 10-best translation hypotheses of this EN-CS NMT model on training and validation instances are re-ordered by sentence-level METEOR scores and then top 4 are labeled positive.

Then, we train multimodal binary classifiers:

1. **Random Forest (SHEF_ARF)**

2. **RNN Classifier (SHEF_ARNN)**

---

**Why cross-lingual WSD?**  
We believe humans usually look at the image to disambiguate ambiguous words in the source sentence and then select correct translation.

**Why Re-ranking?**  
In a preliminary experiment, we trained a standard NMT and looked at 20-best hypotheses. The best of 20-best hypotheses (Oracle) was compared to the 1-best hypothesis and we found plenty of scope for re-ranking these hypotheses.

**Why no difference?**  
For EN-FR, only 12% to 15% of the test instances get re-ranked.

And, less than 2% lexical changes

Contribution of re-ranking is small and since SHEF_LT (Text-only) and SHEF_MLT (Image-aware) outputs are nearly identical, the contribution of image is further minute.

**Task1b** shows a similar trend: the contribution of images is minor. The best model was SHEF_CON – the consensus-based model that does not use image information.