

Movie Translation

Vikram Voleti

December 2017

1 Aim

To translate (English, or other) movie scenes into Indian (or another) language.

2 Modes

The task operates on three modes:

2.1 Text (Dialogues)

The movie's dialogues need to be translated from the original language (English, say) to the target language (Hindi, say). The original dialogues could be obtained from the movie's subtitles (or, speech recognition ([Deep Speech](#), [wit.ai](#), [Google API](#)) could be used).

- [Facebook MT](#) (Convolutional Sequence to Sequence Learning)
- [Unsupervised NMT](#) (Kyunghyun Cho, NYU)
- [Unsupervised Machine Translation Using Monolingual Corpora Only](#) (FAIR-Paris)
- [Fully-parallel text generation for neural machine translation](#) (einstein.ai)

2.2 Speech

Once the target dialogues have been translated, the corresponding speech needs to be generated using the same voice.

- [WaveNet](#) (DeepMind)
- [Deep Voice 1 & 2](#) (Baidu)
- [A study of speaker adaptation for DNN-based speech synthesis](#) (Interspeech, 2017)
- [Voice Conversion from Unaligned Corpora using VAW-GANs](#) (Interspeech, 2017)

2.3 Lips

Next, lip movements need to be warped to model the translated text.

- [You Said That?](#) (Zisserman, BMVC 2017)

- [Lipsyncing Obama](#) (SIGGRAPH 2017), [UWash page](#)
- [A Deep Learning Approach for Generalized Speech Animation](#) (SIGGRAPH 2017)
- [PHOTO-REAL TALKING HEAD WITH DEEP BIDIRECTIONAL LSTM](#) (ICASSP 2015, Microsoft), [A deep bidirectional LSTM approach for video-realistic talking head](#) (Journal, Multimedia Tools and Applications, 2016)
- [Audiovisual speech synthesis: An overview of the state-of-the-art](#) (Journal, Speech Communication, 2015); [Comprehensive many-to-many phoneme-to-viseme mapping](#) and its application for concatenative visual speech synthesis (Journal, Speech Communication, 2013)