Synthetic and Natural Noise Both Break Neural Machine Translation

Character based neural models help us counteract out of vocabulary issues(also reduce the computation of large word embedding matrices) but they perform poorly in case of noisy data.
Authors present approaches to make model robust to external noise.

Current NMT systems perform poorly in case of noisy test input(Assumption of current models is that the training input would also have been noisy and model has learnt to deal with the noise from the training data)

Check paper for character level models experiment details:
Authors test the model on natural noise and also on synthetically generated noise to show that current char level models(charCNN, char2char and Nematus) do not perform well.
They also show that using google translate on Noisy data and then doing prediction correction also gives poor results.
Thus, the conclusion is that there is need for a rich grammatical model to do prediction correction.

(Black box and white box adversarial training)
2 methods used to deal with noise:

1. The previous character level models encode sentences one character at a time. Thus the order of character in the words is considered to learn a representation.
    a. Since noise scrambles(the synthetic type considered in the paper) the order of alphabets in the words, it makes sense to learn a character level representation that does not consider the order of characters in the word
    b. The authors propose to sum the character embeddings of a given word and then use that to learn word level encoder.
    c. However, this does not seem promising as the order of characters in word is lost and naturally the performance is degraded.
2. Train models by externally introducing noise during training.

Authors also hypothesize why charCNN model was able to perform good even on

mixed noise while other models could not.
The reason is that different convolution filters learn to deal with different kinds of noise.

Other conclusions:

- Natural noise is very different from noise synthetically generated in the paper.
- Models performing well on synthertic noise did not do well on natural noise.