Which Settings Give Better Results?

- Experiments on image captioning in Chinese for Flickr8k and Flickr30k images

Xi Chen, University of Gothenburg

Abstract

In this project, an experimental study in image captioning in Chinese language is carried out. Totally twelve models were trained with different settings (such as both applying and not applying Chinese text segmentation, using both pre-trained VGG19 and ResNet101, both with and without fine-tuning, etc.) on both Flickr8k-CN and Flickr30k-CN datasets, and twenty-four checkpoints were examined. The trained models/checkpoints were then evaluated by BLEU, ROUGE_L and CIDEr criterias for comparison and to see how they performed. Evaluation results were shown in tables in the appendix, and discussed. Ten captioning examples are also shown in the appendix.

1. Introduction

Image caption as a significant part of artificial intelligence has been deeply researched by many researchers in the past decade. However, most of the research was based on image captioning in English language, and most of the databases for research that can be found are also English-based. Chinese is the world's most spoken language, but related research in Chinese language in the field of natural language processing and computer vision, especially in image captioning still cannot compete with those in English language. Some researchers have realized this and have done several pilot research on this topic. Inspired by their work, this little project was conducted in order to do some experiments on image captioning in Chinese language.

1.1 Related works

The authors of [1] have extended the popular Flickr8k with both machine-translated and human-annotated Chinese captions and experimented how they performed in image captioning in Chinese, and concluded that "Baidu translation is preferred to Google translation", and "which model is more suited for Chinese captioning depends on what ground truth is used" and the NIC model performed in image captioning in Chinese just as well as in image captioning in English. H. Peng & N. Li [2] have in their experiment showed that according to the BLEU score the char-level method worked better than the word-based method in generating Chinese captions for Flickr30K images. However, the ground truth Chinese captions they used are directly translated from English captions in the Flickr30K database by Google Translation API. In the work [3], the authors has

extended the popular Flickr30k database with machine-translated Chinese captions and then proposed a model which can estimate the fluency of the machine-translated sentences in order to decide if they should be used in the image captioning model, or their importance should be decreased because of low fluency when fed into the image captioning model, which provides us a way to use machine-translated ground truth in image captioning, and at the same time helps to increase the performance of the image captioning model by excluding bad machine-translations.

1.2 Goal

The goal of this project is not to propose any model for image captioning in Chinese which can compete with the state-of-the art researches, but rather, based on the code of [4], conduct several experiments on generating Chinese captions for both Flickr8K and Flickr30K images with ground truth captions from Flickr8K-CN and Flickr30K-CN, using trained models with different settings; and to finally evaluate their performance using BLEU, ROUGH_L and CIDEr scores when different beam size are chosen.

By “models with different settings”, it is meant that models using both char-based method and segmentation-based models, using both pre-trained Resnet101 and VGG19, both without fine-tune and with fine-tune, both with best BLEU-4 scores and with early stop.

1.3 Data

The data used in this project is Flickr8k-CN and Flickr30k-CN. [5] Flickr8k-CN is a dataset of images with Chinese captions extended by [1] from the popular Flickr8k dataset. Flickr30k-CN is a dataset extended from Flickr30k by [3] which contains images with Chinese captions; however, all the captions in the train set and val set are machine-translated, but human-translated in the test set. The splits of these two datasets are shown in the following table.

	Flickr8k-CN			Flickr30k-CN		
	train	val	test	train	val	test
Images	6000	1000	1000	29,783	1000	1000
Human-annotated Chinese sentences	30,000	5000	5000			
Machine-translated Chinese sentences	30,000	5000		148,915	5000	
Human-translated Chinese sentences			5000			5000

2. Methods

The methods used for this project will be explained in the following parts: the code used for programming, the model used for training, the four different settings used for training in order to do the experiments to achieve this project's goals, and the methods used for evaluation of the results.

2.1 Code

The code of this project is mainly based on [6]. It is a tutorial code to image captioning using pytorch [7]. The image captioning model used in this tutorial is based on the paper [8], which to some extent no longer is the state-of-the-art model, but still suitable for pilot studies. The code for this project is available at:
https://github.com/guschenxi/aics-project.

Some files have been added and many files in this tutorial have been rewritten to suit for this project. "create_json_flickr30kcn.ipynb" and "create_json_flickr30kcn.py" were added in order to transfer the data from Flickr30k-CN and Flickr8k-CN into the required form which can be used as the input data into the model. Minor changes have been made in "model.py" and "train.py" in order to make them fit the experiment settings. Greedy search and beam search functions have been added into "utils.py". "checkpoints.py" was added to store all the names of the trained models with different settings. "eval.py" was re-written so that it was able to conduct evaluation for all models with different settings, with beam size from 1 to 5. "caption_greedy.ipynb" and "caption_beam.ipynb" were added to perform image captioning in Chinese for raw images using greedy search and beam search, and to illustrate the captioning results together with the attention distribution.

2.2 Model

The model for this project is based on the paper [8]. It consists of an encoder and decoder with an attention mechanism which is a typical structure of a model for image captioning. The encoder takes in image inputs into pre-trained Convolutional Neural Networks and encodes them into smaller representations of learned important features from the images.

Since image captions are tokens with sequence, the decoder is therefore an Recurrent Neural Network, here in this project, an LSTM (Long Short-Term Memory) which takes a look at the encoded images and generates a caption of the image word by word.

The attention mechanism enables the encoder to look at the corresponding part of the image when generating each token, i.e. the pixels of the more important part of the image corresponding to the token get bigger weights.

2.3 Experiment Settings

2.3.1 Chinese Text Segmentation

Different to English and most other languages, Chinese is a character-based language rather than an alphabet and word-based language. Moreover there are no spaces between characters (or character groups which can have a separate semantic meaning). Therefore, tokenization for Chinese language processing is not as simple as in English which can be done by splitting a sentence by spaces. However, there are some popular tools which provide Chinese text segmentation, such as “Jieba” [9]. The Flickr8k-CN and Flickr30k-CN databases used in this project were already tokenized, so that there are added spaces between each semantic unit (either character or character group). In this project, both character-based models (by taking away the added spaces, and splitting the sentence again by characters) and token-based (segmentation-based) models were trained and experimented, to see how well they perform.

2.3.2 ResNet101 and VGG19

With the rapid development of computer vision, more researchers have presented deeper and deeper neural networks in order to deal with more complex image-related problems. VGG and ResNet were two of those presented deeper neural networks which are widely used nowadays. [10]

VGG (stands for Visual Geometry Group, a group of researchers at Oxford who developed this architecture [11]) is a convolutional neural network structure presented in 2014 by Simonyan and Zisserman. [12] According to their paper, the model achieves 92.7% top-5 test accuracy in ImageNet. VGG19 is a 19 layer deep variant of VGG networks.

ResNet, short form of Residual Network is another neural network structure presented in 2015 by K. He et al. [13] ResNet has similar structure as VGG, but can better deal with the problem of vanishing gradients. ResNet has better performance which achieves 93.3% top-5 test accuracy on ImageNet, than VGG, and works faster than VGG. ResNet101 is a Residual Network variant which is 101 layer deep.

Pre-trained models that are trained on large databases and contain feature representations of the data they are trained on can be used to different data in order to save a great amount of time. The learned features are transferable and can benefit other models. The pre-trained models of VGG19 and ResNet101 are provided by most deep learning APIs, as well as Pytorch. In this project, both pre-trained VGG19 and pre-trained ResNet101 were experimented in order to examine their performance.

2.3.3 Feature Extraction and Fine Tune

Using pre-trained models which contain transferable learned features to train new models is considered as transfer learning. This brings out two approaches to transfer

learning: feature extraction and fine tuning.[14] Feature extraction means that the parameters of the pre-trained model are not changeable, and the feature representations learned will be used directly in the new model; while fine tuning allows the weights in the pre-trained model to be changed and fine-tuned, so that it will be adapted to and benefits the new task. In order to see how allowing fine tuning will affect the performance of the models, experiments both without fine tuning (feature extraction) and with fine tuning were conducted in this project.

2.3.4 Models with and without Best BLEU Scores

BLEU (bilingual evaluation understudy) is an approach brought up by Kishore Papineni, et al. in 2012, which originally was for evaluating the quality of machine translation. The BLEU score is to calculate how many n-grams in the candidate translation match the n-grams in the reference translation. The calculated BLEU score is a number between 0 and 1, which shows how close a machine translation is to a professional human translation. [15] BLEU-1, BLEU-2, BLEU-3 and BLEU-4 refer to the cumulative 1-gram, 2-gram, 3-gram and 4-gram BLEU scores. Even though BLEU was brought up to evaluate machine translation, it is nowadays broadly used to evaluate image captioning.[adding chinese captions to images] Therefore, BLEU scores were also used in this project to evaluate the performance of different models.

The tutorial code provided by [4] has a setting which saves the checkpoint after each epoch, and replaces it with a new checkpoint after a new epoch is finished. Therefore, the checkpoint after the last epoch is always saved. It also calculates the BLEU-4 score after each epoch, and if the BLEU-4 score has increased compared with it from the previous epoch, the checkpoint will be saved (or replace the previous one) and named with “BEST”. In the tutorial code, it has a mechanism which enables the training process with early stop when the BLEU score doesn't increase after 20 epochs. When training stops, the last checkpoint will replace the previous one and will be used as one of the final models for this project. The 20th checkpoint before this last checkpoint will of course be the one with the best BLEU-4 score and be saved and named with “BEST”, which will also be used as one of the final models. Each of the settings mentioned above got two checkpoints after training, one with best BLEU-4 score and one achieved 20 epochs after the one with best BLEU-4 score.

2.4 Training

Training epochs were set to 100, however, early stopping were triggered in all training processes. All of them early-stopped at around 30-40 epochs. The learning rate for the encoder was set to 1e-4, only when fine tuning was turned on. The learning rate for the decoder was set to 4e-4. If the BLEU score of the validation process after each epoch has not improved in 8 epochs, the learning rate will be decreased by 0.8. The dimension of word embeddings and of the attention layers was set to 512. The hyperparameters used for training were the same as the tutorial code, remaining not changed. The only difference is that when conducting

experiments on using pre-trained VGG19, the dimensions have been changed to adapt VGG19's structure. The whole training process will not be explained in detail here, since no significant change has been made compared to the original tutorial.

After training using 6 different experiment settings shown in the following table, on both Flickr8k-CN and Flickr30k-CN, saving two checkpoints (explained above in 2.3.4) for each, totally 12 training sessions with 6 different settings were conducted and 24 checkpoints were created.

	Character-based	Segmenta-tion-based	Pre-trained VGG19	Pre-trained ResNet101	Feature Extraction	Fine-tuning
Setting 1	x			x	x	
Setting 2		x		x	x	
Setting 3	x			x		x
Setting 4		x		x		x
Setting 5	x		x		x	
Setting 6		x	x		x	

2.5 Evaluation

The criterias used for evaluation are BLEU, ROUGE_L [16, 17] and CIDEr [18], same as many research have done. [1,2,3] Meteor is not possible to apply, since the Chinese language is a character-based language, the notion stemming does not apply, and a dictionary for synonyms in Chinese lacks.

From the beginning, the BLEU scores were calculated using the "corpus_bleu" function inside nltk.translate toolkit [19]. Then an evaluation code for natural language generation which automatically conducts evaluation with different criterias was found at [20]. However, some minor changes have been made to fit this project and some small bugs in the code were also fixed.

Beam search algorithm [6, 21] is used in this project for generating image captions and then for evaluation. It means that at each time step, the n most possible tokens are chosen for predicting the next token. In this way, a tree of n^n branches will be created and each branch represents a set sequenced tokens. Accumulated scores will be calculated for all branches, and then the branch with the highest score will be the final sentence. When the beam size n equals 1, the algorithm is the same as the greedy search algorithm, which means that at each time step, only the token with the highest possibility will be used for generating the next token. The problem with greedy search is

that even though the first token has the best score, the followed tokens decided by the first token may not be the best choices (have high scores); while with beam search algorithm, the most optimal result will not be decided until all sequences of tokens are generated and their scores are calculated. In this project, the beam size n is tested from 1 to 5 to how it affects the final results.

3. Evaluation Results

The results of the evaluation of the trained models are shown in tables in the appendix attached below this report. Some of the experiment settings are written in short forms. “BB” refers to the models with **B**est **B**LEU scores as described in section 2.3.4; while “NB” (**N**ot **B**est BLEU) refers to the models (checkpoints) trained 20 epochs more than “BB” models. “8k” and “30k” refers to models trained on Flickr**8k**-CN and Flickr**30k**-CN. “FE” and “FT” refers to models trained without fine-tuning (**F**eature **E**xtraction) and with **F**ine-**T**uning. “RN” and “VG” refers to models trained using pre-trained **R**es**N**et101 and **VG**G19. “b1”-“b5” means evaluation results using different **b**eam sizes from 1 to 5. “B-1”, “B-2”, “B-3”, “B-4” refers to **B**LEU**-1**, **B**LEU**-2**, **BL**EU**-3** and **B**LEU**-4** scores.

Generally, it can be seen that models trained on Flickr8k-CN got much better scores than models trained on Flickr30k-CN. One possible reason is that the ground truth captions in the train and val split of Flickr30k-CN were machine-translated, some of which lack fluency.[3] In the following of this chapter, evaluation results are shown depending on the choices of seam size and the settings of the models.

3.1 Which Beam Size Performs Better?

The results (Result 1A & 1B) do not show a fixed pattern about which beam size generates sentences with higher scores. The performance depends on the settings of the model, i.e. which dataset it used, if fine-tuning was turned on or not, which pre-trained model was used in the encoder, etc. However, in most cases, beam size between 3 and 5 gave higher scores.

3.2 Character-Based vs Segmentation-Based

As a native Chinese speaker, I thought that segmentation-based models should work much better than character-based, since segmentation is as words in English, and it is how the language should “work”. To my surprise, as can be seen in the whole table (Result 2), except four exception scores which occurred when beam size equals 1, almost all character-based models have gotten higher scores than those with exactly the same settings but segmentation-based models. This also proves [2]’s previous finding.

3.3 ResNet101 vs VGG19

As the results (Result 3) in the table shows, almost all models using pre-trained ResNet101 have achieved better BLEU, ROUGE_L and CIDEr scores than those with exactly the same settings but using pre-trained VGG19. However, one exception is the model setting marked as “BB-30k-seg-FE”, in which almost all BLEU and ROUGE_L scores from the model using pre-trained VGG19 outride that using pre-trained ResNet101. Yet, the CIDEr score still shows unchanged results.

3.4 Feature Extraction vs Fine-Tuning

The results (Result 4) in the first table indicate that, for almost all of the models marked by “BB”, fine-tuning the encoder gave better BLEU and ROUGE_L scores, than only using the encoder for feature extraction. On the contrast, the situation changed when it came to models marked by “NB” (models trained for 20 more epochs): not fine-tuning the encoder gave better BLEU and ROUGE_L scores for most models than those using fine-tuning. When it comes to CIDEr score, almost all of the models without fine-tuning got better CIDEr scores, except one marked as “BB-30k-seg-RN”

3.5 Checkpoints with Best BLEU Score vs Continue Training (20 More Epochs)

(Result 5) A large percent of the scores (with several exceptions) show that models marked by “BB” achieved higher scores than those with exactly the same setting but marked by “NB”, which means that most of the models trained for 20 more epochs do not perform better than the checkpoints created 20 epochs earlier. It can be concluded that models trained for more epochs do not always perform better, but can lead to overfitting and therefore worse performance. Early-stopping and saving it as the “BEST” when BLEU-4 score does not increase more, as the tutorial code programmed, seems to be a good approach. However, other criterias other than BLEU remain to be tested in this context.

Some data has shown that when beam size = 1 (greedy search), some models trained for 20 more epochs could achieve better scores, but not always; when beam size > 1, the models trained for 20 less epochs always achieved better scores.

4. Discussion

4.1 Evaluation Criterias

To some extent, describing an image is subjective. Describing an image in a specific language depends on even more, such as semantic and syntactic aspects of the language, culture background, etc. As many researchers have pointed out, evaluating a generated image captions just by some criterias is not enough and not always reliable. [22] It has happened that some BLEU scores or ROUGE_L scores in the results showed that one model is better than the compared model, while the CIDEr showed the opposite result. Human evaluation has also been used in many researches. However, due to limited time and resources, evaluation by only criterias was carried out, and it was enough to discover some patterns in the scores.

4.2 Training Data Bias

The train set and val set in Flickr30k-CN consists of machine-translated Chinese sentences from English, while the test set consists of human-translated sentences,

which was probably the reason why models trained on Flickr30k-CN got much lower scores, compared with those trained on Flickr8k-CN which consists of only human-translated and human-annotated sentences.

As we can see in Captioning Examples 3, 7 and 9, the pictures have nothing to do with humans or animals. One of them shows a basket of fruits and the other one shows a computer screen. But the generated results seemed not good. There was always a person or even more than one (a man and a woman) existing in the captions. This is probably because most of the captions in the train set contain people or animals, which made the models learn that the possibilities of these tokens describing people or animals appearing are extremely high.

When I fed the models with a picture of zebras, they generated “dogs” and “birds” instead. This error also depends on the train set. The tokens for zebra probably do not exist in the training data. This also happened when a bear was in the picture, but recognized as a dog.

4.3 Gender Bias

When I tried to feed the models with a picture in which a man is holding a baby, many of the models generated sentences like “a woman is holding a boy”, even though the features of the man were very clear.

When I fed the models with a picture in which a man in red shirt is standing on the street, some models generated sentences like “a woman in red dress is walking on the street”, while some generated “a person in red clothes is walking on the street”, while others generated “a man in red shirt is on the street”.

The gender bias has been widely discussed by other researchers. It has undoubtedly happened in this project, since the model used for this project was just a basic model.

4.4 Quantity Errors

Quantity errors occurred often when testing generating captions. Here are some sentences generated for the picture on the left by some of the models. It shows that sometimes the quantity word describing people are wrong, while sometimes the quantity word describing motorcycles are wrong. However, there are also correctly generated sentences. This probably also depends on the dataset for training. The probability of the forms of the words appearing in the training data decides how the models count objects.

['<start>', '一个', '骑', '摩托车', '的', '人', '骑', '着', '摩托车', '<end>']
A person who’s riding **a** motorbike riding **a** motorbike
['<start>', '一', '个', '骑', '着', '摩', '托', '车', '的', '警', '察', '<end>']
A policeman/woman riding **a** motorbike

['<start>', '一', '群', '人', '在', '一', '辆', '摩托车', '上', '<end>']
A group of people on **one** motorbike
['<start>', '一', '群', '骑', '摩', '托', '车', '的', '人', '<end>']
A group of people ride motorbikes
['<start>', '一', '群', '人', '在', '摩托车', '上', '<end>']
A group of people on motorbikes
['<start>', '一', '群', '人', '骑', '着', '摩', '托', '车', '<end>']
A group of people riding motorbikes
['<start>', '骑', '摩', '托', '车', '的', '人', '<end>']
People on motorbike (without singular or plural forms)
['<start>', '一个', '骑', '摩托车', '的', '人', '<end>']
A person riding **a** motorbike
['<start>', '骑', '自', '行', '车', '的', '人', '<end>']
Person riding **a bike**
['<start>', '一', '群', '人', '骑', '着', '一', '辆', '摩托车', '<end>']
A group of people riding **one** motorbike

4.5 Segmentation-Based and Character-Based

As Captioning Example 10 shows: the model marked as “BB-30k-seg-FT-RN” could not correctly generate the token (grouped characters) for “frisbee”, while the corresponding “char” model generated correctly the token (two separate characters). This is because that when using Chinese text segmentation, several grouped characters are encoded as one token, the chance that the token, in this case “飞盘 (frisbee)”, appears in the whole vocabulary has dropped to smaller than 5 and therefore will be replaced by “<unk>”. On the contrary, when using character-based splitting, “飞 (fly)” and “盘 (plate)” will be used as two different tokens and the chance they appear in the vocabulary has strongly increased. The character-based models can still learn the possibility of these two words appearing together, in order to generate correct tokens. This explains why character-based models got higher scores than segmentation-based models.

4.6 Interrupted Training

An experience when training these models was that: several training sessions were terminated in the middle of the training, because of bad Internet connection. Then they were trained continually from the epoch at which they were interrupted. Some of their batch sizes were changed from 32 to 16 when resuming the training processes, because of GPU’s capability at those moments. However, when their evaluation results were compared then, some of the scores seemed very strange compared to other models. Then they were trained from the beginning and evaluated again. The final scores were totally different and varied to a big extent, so they got corrected at last. Even though the tutorial code provides with the function which enables the checkpoints to be saved and restored from the middle, it seemed according to the evaluation scores that errors have occurred anyway which affected the final comparison among the

models. In future works, interrupted training sessions should be re-trained from beginning rather than restored from saved checkpoints.

5. Conclusions and Further Work

In this project, based on existing research and programming code, an experimental study has been carried out, to examine how different settings for model training and different beam size for generating captions can affect the performance in image captioning in Chinese Language, by evaluating their BLEU, ROUGE_L and CIDEr scores. After comparing the scores, several findings were concluded and discussed.

In future works, training the models on another dataset, such as MS COCO, should be carried out in order to examine how different datasets and data quality affect the results. Gender bias and quantity errors should be examined and optimized. Evaluation with other standards and even by humans rather than BLEU, ROUGH_L and CIDEr is also worth implementing in future works. Image captioning in Chinese is still a field lacking research and remains to be improved.

References

[1] X. Li, W. Lan, J. Dong, and H. Liu, “Adding Chinese captions to images,” in ICMR, 2016.
[2] H. Peng and N. Li, “Generating Chinese Captions for Flickr30K Images,” 2016
[3] W. Lan, X. Li, and J. Dong, “Fluency-Guided Cross-Lingual Image Captioning,” in ACM on Multimedia Conference (ACMM), 2017, pp. 1549–1557
[4] https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Image-Captioning
[5] https://github.com/li-xirong/cross-lingual-cap
[6] https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Image-Captioning
[7] https://pytorch.org/
[8] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation with visual attention. arXiv preprint arXiv:1502.03044 (2015)
[9] https://github.com/fxsjy/jieba
[10] A. Anwar, “Difference between AlexNet, VGGNet, ResNet, and Inception”, retrieved on 2021-02-15 from https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96,
[11] S. Sarin, “VGGNet vs ResNet”, retrieved on 2021-02-15 from https://towardsdatascience.com/vggnet-vs-resnet-924e9573ca5c
[12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in International Conference on Learning Representations (ICLR), 2015.
[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv:1512.03385, 2015.
[14] Matthew Peters, Sebastian Ruder, and Noah A Smith. “To tune or not to tune? Adapting pretrained representations to diverse tasks.” 2019. ArXiv:1903.05987.

[15] https://en.wikipedia.org/wiki/BLEU
[16] “What is ROUGE and how it works for evaluation of summaries?”, retrieved from http://text-analytics101.rxnlp.com/2017/01/how-rouge-works-for-evaluation-of.html, on 2021-02-15
[17] Lin, Chin-Yew. 2004. ROUGE: a Package for Automatic Evaluation of Summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain, July 25 - 26, 2004.
[18] R. Vedantam, C. L. Zitnick, and D. Parikh. CIDEr: Consensus-based image description evaluation. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015
[19] https://www.nltk.org/api/nltk.translate.html
[20] https://github.com/Maluuba/nlg-eval
[21] https://en.wikipedia.org/wiki/Beam_search
[22] Rachael Tatman, “Evaluating Text Output in NLP: BLEU at your own risk.” Retrieved at https://towardsdatascience.com/evaluating-text-output-in-nlp-bleu-at-your-own-risk-e8609665a213, on 2021-02-17

Appendix

Result 1-A						3.1 Which Beam Size Performs Better?					
						B-1	B-2	B-3	B-4	ROUGE_L	CIDEr
BB	30k	char	FE	RN	b1	**31.99**	20.39	12.75	8.34	29.63	39.79
					b2	30.26	19.82	12.80	8.66	30.66	45.33
					b3	30.86	20.22	13.10	8.90	30.73	46.10
					b4	31.22	**20.55**	**13.35**	**9.10**	**30.95**	**46.87**
					b5	31.18	20.49	13.31	9.06	30.87	46.56
BB	30k	char	FT	RN	b1	30.91	19.74	12.22	7.92	29.43	37.99
					b2	31.16	20.67	13.39	9.06	31.09	**46.52**
					b3	31.53	21.02	13.75	9.39	31.16	46.11
					b4	31.51	21.06	13.86	9.53	31.25	46.20
					b5	**31.53**	**21.11**	**13.92**	**9.58**	**31.26**	46.23
BB	30k	seg	FE	RN	b1	**26.66**	**12.86**	7.09	4.05	24.01	29.18
					b2	22.97	12.27	7.26	4.43	24.99	37.10
					b3	23.38	12.68	7.61	4.68	25.20	37.37
					b4	23.51	12.79	**7.71**	4.75	**25.36**	**37.63**
					b5	23.47	12.75	7.68	**4.74**	25.35	37.45
BB	30k	seg	FT	RN	b1	**26.62**	13.32	7.45	4.33	25.35	32.73
					b2	23.54	12.80	7.67	4.76	25.77	37.15
					b3	24.34	13.42	8.10	4.99	**26.28**	38.17
					b4	24.28	**13.46**	8.15	**5.07**	26.26	38.10
					b5	24.24	13.44	**8.16**	5.06	26.26	**38.25**
BB	8k	char	FE	RN	b1	63.01	49.63	38.40	29.55	49.56	59.91
					b2	67.05	54.35	43.20	34.20	50.89	63.77
					b3	**67.67**	**54.93**	**43.59**	**34.54**	**51.21**	**64.89**
					b4	67.62	54.83	43.50	34.47	51.04	64.54
					b5	67.49	54.61	43.20	34.13	50.89	64.15
BB	8k	char	FT	RN	b1	59.05	45.72	34.50	25.90	48.34	53.59
					b2	**68.23**	55.22	43.75	34.61	51.03	62.17
					b3	68.22	**55.41**	44.10	34.99	**51.43**	**63.75**
					b4	68.19	55.29	43.95	34.88	51.33	63.27
					b5	68.21	55.38	**44.12**	**35.08**	51.39	63.54
BB	8k	seg	FE	RN	b1	61.76	43.30	30.06	20.93	46.13	49.13
					b2	62.45	46.66	34.56	25.53	46.68	53.35
					b3	63.10	46.90	34.72	25.58	46.67	52.32
					b4	63.18	47.17	35.15	26.11	47.02	53.29
					b5	**63.32**	**47.27**	**35.23**	**26.21**	**47.06**	**53.66**
BB	8k	seg	FT	RN	b1	**65.34**	46.03	32.10	22.57	47.15	49.10
					b2	64.16	48.05	35.52	26.23	47.64	51.63
					b3	65.00	**48.80**	**36.10**	**26.69**	47.83	51.74
					b4	65.00	48.71	36.02	26.56	**48.06**	**52.38**
					b5	64.98	48.63	35.91	26.47	47.94	52.06
BB	30k	char	FE	VG	b1	30.37	19.24	11.88	7.65	28.99	38.22
					b2	29.31	19.20	12.39	8.32	30.31	43.84
					b3	29.58	19.42	12.55	8.45	30.42	44.30
					b4	29.72	19.53	12.65	8.53	30.45	44.33
					b5	**29.75**	**19.54**	**12.66**	**8.54**	**30.50**	**44.49**
BB	30k	seg	FE	VG	b1	25.50	12.43	6.89	3.89	24.07	28.68
					b2	23.57	12.74	7.49	4.46	24.77	34.26
					b3	24.30	13.39	7.95	4.79	25.30	36.22
					b4	24.67	13.66	8.17	4.97	25.51	**36.94**
					b5	**24.72**	**13.69**	**8.19**	**4.98**	**25.53**	36.87
BB	8k	char	FE	VG	b1	56.25	43.49	32.84	24.67	47.77	54.29
					b2	66.36	**53.22**	**41.74**	**32.52**	50.24	60.30
					b3	**66.44**	53.12	41.46	32.29	**50.24**	60.34
					b4	66.41	53.13	41.51	32.37	50.12	60.74
					b5	66.22	52.97	41.37	32.27	50.13	**60.80**
BB	8k	seg	FE	VG	b1	57.35	39.32	26.51	18.10	45.02	44.00
					b2	**63.35**	46.26	33.40	24.21	46.25	49.30
					b3	63.00	46.02	33.41	24.32	46.45	49.82
					b4	63.19	46.23	33.56	24.45	46.52	**49.94**
					b5	63.28	**46.35**	**33.70**	**24.58**	**46.60**	49.94

Result 1-B						3.1 Which Beam Size Performs Better?					
						B-1	B-2	B-3	B-4	ROUGE_L	CIDEr
NB	30k	char	FE	RN	b1	**32.05**	**20.42**	12.72	8.31	29.40	37.31
					b2	29.53	19.39	12.48	8.40	30.56	45.08
					b3	30.47	19.99	12.84	8.66	30.68	45.35
					b4	30.66	20.19	13.06	8.87	30.82	45.78
					b5	30.78	20.26	**13.13**	**8.92**	**30.86**	**45.79**
NB	30k	char	FT	RN	b1	**33.18**	**21.36**	**13.36**	**8.70**	29.63	38.97
					b2	29.19	19.12	12.22	8.15	29.96	**43.19**
					b3	29.87	19.54	12.49	8.31	30.06	42.87
					b4	29.89	19.55	12.52	8.34	30.10	42.62
					b5	29.91	19.57	12.52	8.34	**30.15**	42.71
NB	30k	seg	FE	RN	b1	**26.53**	**13.04**	7.18	4.15	24.22	29.91
					b2	23.22	12.32	7.24	4.36	25.04	35.99
					b3	23.99	12.88	7.60	4.58	25.46	36.67
					b4	24.03	12.88	7.60	4.58	**25.48**	**37.13**
					b5	24.05	12.88	**7.61**	**4.59**	25.45	37.02
NB	30k	seg	FT	RN	b1	**26.98**	**13.40**	**7.35**	4.13	24.66	30.01
					b2	23.11	11.98	6.91	4.13	24.51	34.03
					b3	23.64	12.40	7.22	4.31	24.80	34.43
					b4	23.54	12.36	7.22	4.31	24.78	34.29
					b5	23.54	12.37	7.24	**4.34**	**24.80**	**34.59**
NB	8k	char	FE	RN	b1	61.67	47.53	35.75	26.79	47.33	54.27
					b2	66.01	52.75	41.09	31.85	49.43	60.83
					b3	66.15	53.00	41.35	32.04	**49.96**	61.37
					b4	**66.39**	**53.26**	**41.62**	**32.28**	49.91	**61.40**
					b5	66.17	53.12	41.55	32.25	49.86	61.21
NB	8k	char	FT	RN	b1	59.06	44.92	33.27	24.46	46.35	46.78
					b2	64.49	50.43	38.40	29.11	47.64	51.98
					b3	64.60	50.60	38.57	29.27	47.97	52.54
					b4	**64.76**	**50.75**	**38.74**	**29.46**	**48.09**	**52.74**
					b5	64.71	50.67	38.67	29.38	47.99	52.41
NB	8k	seg	FE	RN	b1	57.91	39.50	26.76	17.99	44.54	44.08
					b2	62.78	44.64	31.43	22.03	45.72	49.23
					b3	62.67	**44.73**	**31.58**	**22.38**	46.01	50.01
					b4	62.66	44.68	31.46	22.20	46.03	50.02
					b5	**62.68**	44.69	31.50	22.28	**46.12**	**50.27**
NB	8k	seg	FT	RN	b1	59.98	40.55	27.79	19.39	44.32	41.74
					b2	62.08	44.14	31.49	22.84	45.25	45.72
					b3	62.42	44.48	31.71	23.01	45.71	46.75
					b4	**62.49**	44.49	31.71	23.00	45.85	47.06
					b5	62.43	**44.49**	**31.76**	**23.06**	**45.90**	**47.18**
NB	30k	char	FE	VG	b1	**31.93**	**20.10**	**12.35**	7.97	28.91	35.75
					b2	29.34	18.81	11.87	7.84	29.69	42.32
					b3	29.78	19.14	12.08	8.00	29.83	43.06
					b4	29.84	19.19	12.12	**8.01**	29.82	43.11
					b5	29.80	19.16	12.10	8.00	**29.85**	**43.61**
NB	30k	seg	FE	VG	b1	**26.90**	**13.07**	7.17	4.02	24.13	28.79
					b2	23.46	12.29	6.98	4.07	24.50	33.74
					b3	23.79	12.59	7.23	4.25	24.76	34.37
					b4	24.06	12.73	**7.29**	**4.28**	24.89	**34.45**
					b5	24.06	12.72	7.28	4.26	**24.89**	34.41
NB	8k	char	FE	VG	b1	60.11	45.61	33.82	25.04	47.01	48.77
					b2	**65.23**	**51.27**	**39.32**	**29.96**	**48.88**	55.85
					b3	65.06	50.82	38.87	29.65	48.59	55.08
					b4	65.06	50.94	39.02	29.81	48.81	55.86
					b5	65.16	50.98	39.04	29.82	48.83	**55.96**
NB	8k	seg	FE	VG	b1	56.97	38.73	26.37	17.95	43.51	41.99
					b2	60.97	42.97	30.24	21.24	44.45	45.89
					b3	61.03	43.28	30.57	21.61	44.38	46.14
					b4	**61.16**	**43.29**	**30.61**	**21.72**	**44.46**	**46.23**
					b5	61.04	43.21	30.58	21.70	44.44	46.00

Result 2 **3.2 Character-Based vs Segmentation-Based**

		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		char	seg	char	seg	char	seg	char	seg	char	seg	char	seg
BB 30k FE RN	b1	**31.99**	26.66	**20.39**	12.86	**12.75**	7.09	**8.34**	4.05	**29.63**	24.01	**39.79**	29.18
	b2	**30.26**	22.97	**19.82**	12.27	**12.80**	7.26	**8.66**	4.43	**30.66**	24.99	**45.33**	37.10
	b3	**30.86**	23.38	**20.22**	12.68	**13.10**	7.61	**8.90**	4.68	**30.73**	25.20	**46.10**	37.37
	b4	**31.22**	23.51	**20.55**	12.79	**13.35**	7.71	**9.10**	4.75	**30.95**	25.36	**46.87**	37.63
	b5	**31.18**	23.47	**20.49**	12.75	**13.31**	7.68	**9.06**	4.74	**30.87**	25.35	**46.56**	37.45
BB 30k FT RN	b1	**30.91**	26.62	**19.74**	13.32	**12.22**	7.45	**7.92**	4.33	**29.43**	25.35	**37.99**	32.73
	b2	**31.16**	23.54	**20.67**	12.80	**13.39**	7.67	**9.06**	4.76	**31.09**	25.77	**46.52**	37.15
	b3	**31.53**	24.34	**21.02**	13.42	**13.75**	8.10	**9.39**	4.99	**31.16**	26.28	**46.11**	38.17
	b4	**31.51**	24.28	**21.06**	13.46	**13.86**	8.15	**9.53**	5.07	**31.25**	26.26	**46.20**	38.10
	b5	**31.53**	24.24	**21.11**	13.44	**13.92**	8.16	**9.58**	5.06	**31.26**	26.26	**46.23**	38.25
BB 30k FE VG	b1	**30.37**	25.50	**19.24**	12.43	**11.88**	6.89	**7.65**	3.89	**28.99**	24.07	**38.22**	28.68
	b2	**29.31**	23.57	**19.20**	12.74	**12.39**	7.49	**8.32**	4.46	**30.31**	24.77	**43.84**	34.26
	b3	**29.58**	24.30	**19.42**	13.39	**12.55**	7.95	**8.45**	4.79	**30.42**	25.30	**44.30**	36.22
	b4	**29.72**	24.67	**19.53**	13.66	**12.65**	8.17	**8.53**	4.97	**30.45**	25.51	**44.33**	36.94
	b5	**29.75**	24.72	**19.54**	13.69	**12.66**	8.19	**8.54**	4.98	**30.50**	25.53	**44.49**	36.87
BB 8k FE RN	b1	**63.01**	61.76	**49.63**	43.30	**38.40**	30.06	**29.55**	20.93	**49.56**	46.13	**59.91**	49.13
	b2	**67.05**	62.45	**54.35**	46.66	**43.20**	34.56	**34.20**	25.53	**50.89**	46.68	**63.77**	53.35
	b3	**67.67**	63.10	**54.93**	46.90	**43.59**	34.72	**34.54**	25.58	**51.21**	46.67	**64.89**	52.32
	b4	**67.62**	63.18	**54.83**	47.17	**43.50**	35.15	**34.47**	26.11	**51.04**	47.02	**64.54**	53.29
	b5	**67.49**	63.32	**54.61**	47.27	**43.20**	35.23	**34.13**	26.21	**50.89**	47.06	**64.15**	53.66
BB 8k FT RN	b1	59.05	**65.34**	45.72	**46.03**	**34.50**	32.10	**25.90**	22.57	**48.34**	47.15	**53.59**	49.10
	b2	**68.23**	64.16	**55.22**	48.05	**43.75**	35.52	**34.61**	26.23	**51.03**	47.64	**62.17**	51.63
	b3	**68.22**	65.00	**55.41**	48.80	**44.10**	36.10	**34.99**	26.69	**51.43**	47.83	**63.75**	51.74
	b4	**68.19**	65.00	**55.29**	48.71	**43.95**	36.02	**34.88**	26.56	**51.33**	48.06	**63.27**	52.38
	b5	**68.21**	64.98	**55.38**	48.63	**44.12**	35.91	**35.08**	26.47	**51.39**	47.94	**63.54**	52.06
BB 8k FE VG	b1	56.25	**57.35**	**43.49**	39.32	**32.84**	26.51	**24.67**	18.10	**47.77**	45.02	**54.29**	44.00
	b2	**66.36**	63.35	**53.22**	46.26	**41.74**	33.40	**32.52**	24.21	**50.24**	46.25	**60.30**	49.30
	b3	**66.44**	63.00	**53.12**	46.02	**41.46**	33.41	**32.29**	24.32	**50.24**	46.45	**60.34**	49.82
	b4	**66.41**	63.19	**53.13**	46.23	**41.51**	33.56	**32.37**	24.45	**50.12**	46.52	**60.74**	49.94
	b5	**66.22**	63.28	**52.97**	46.35	**41.37**	33.70	**32.27**	24.58	**50.13**	46.60	**60.80**	49.94

		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		char	seg	char	seg	char	seg	char	seg	char	seg	char	seg
NB 30k FE RN	b1	**32.05**	26.53	**20.42**	13.04	**12.72**	7.18	**8.31**	4.15	**29.40**	24.22	**37.31**	29.91
		29.53	23.22	**19.39**	12.32	**12.48**	7.24	**8.40**	4.36	**30.56**	25.04	**45.08**	35.99
		30.47	23.99	**19.99**	12.88	**12.84**	7.60	**8.66**	4.58	**30.68**	25.46	**45.35**	36.67
		30.66	24.03	**20.19**	12.88	**13.06**	7.60	**8.87**	4.58	**30.82**	25.48	**45.78**	37.13
		30.78	24.05	**20.26**	12.88	**13.13**	7.61	**8.92**	4.59	**30.86**	25.45	**45.79**	37.02
NB 30k FT RN	b1	**33.18**	26.98	**21.36**	13.40	**13.36**	7.35	**8.70**	4.13	**29.63**	24.66	**38.97**	30.01
		29.19	23.11	**19.12**	11.98	**12.22**	6.91	**8.15**	4.13	**29.96**	24.51	**43.19**	34.03
		29.87	23.64	**19.54**	12.40	**12.49**	7.22	**8.31**	4.31	**30.06**	24.80	**42.87**	34.43
		29.89	23.54	**19.55**	12.36	**12.52**	7.22	**8.34**	4.31	**30.10**	24.78	**42.62**	34.29
		29.91	23.54	**19.57**	12.37	**12.52**	7.24	**8.34**	4.34	**30.15**	24.80	**42.71**	34.59
NB 30k FE VG	b1	**31.93**	26.90	**20.10**	13.07	**12.35**	7.17	**7.97**	4.02	**28.91**	24.13	**35.75**	28.79
		29.34	23.46	**18.81**	12.29	**11.87**	6.98	**7.84**	4.07	**29.69**	24.50	**42.32**	33.74
		29.78	23.79	**19.14**	12.59	**12.08**	7.23	**8.00**	4.25	**29.83**	24.76	**43.06**	34.37
		29.84	24.06	**19.19**	12.73	**12.12**	7.29	**8.01**	4.28	**29.82**	24.89	**43.11**	34.45
		29.80	24.06	**19.16**	12.72	**12.10**	7.28	**8.00**	4.26	**29.85**	24.89	**43.61**	34.41
NB 8k FE RN	b1	**61.67**	57.91	**47.53**	39.50	**35.75**	26.76	**26.79**	17.99	**47.33**	44.54	**54.27**	44.08
		66.01	62.78	**52.75**	44.64	**41.09**	31.43	**31.85**	22.03	**49.43**	45.72	**60.83**	49.23
		66.15	62.67	**53.00**	44.73	**41.35**	31.58	**32.04**	22.38	**49.96**	46.01	**61.37**	50.01
		66.39	62.66	**53.26**	44.68	**41.62**	31.46	**32.28**	22.20	**49.91**	46.03	**61.40**	50.02
		66.17	62.68	**53.12**	44.69	**41.55**	31.50	**32.25**	22.28	**49.86**	46.12	**61.21**	50.27
NB 8k FT RN	b1	59.06	**59.98**	**44.92**	40.55	**33.27**	27.79	**24.46**	19.39	**46.35**	44.32	**46.78**	41.74
		64.49	62.08	**50.43**	44.14	**38.40**	31.49	**29.11**	22.84	**47.64**	45.25	**51.98**	45.72
		64.60	62.42	**50.60**	44.48	**38.57**	31.71	**29.27**	23.01	**47.97**	45.71	**52.54**	46.75
		64.76	62.49	**50.75**	44.49	**38.74**	31.71	**29.46**	23.00	**48.09**	45.85	**52.74**	47.06
		64.71	62.43	**50.67**	44.49	**38.67**	31.76	**29.38**	23.06	**47.99**	45.90	**52.41**	47.18
NB 8k FE VG	b1	**60.11**	56.97	**45.61**	38.73	**33.82**	26.37	**25.04**	17.95	**47.01**	43.51	**48.77**	41.99
		65.23	60.97	**51.27**	42.97	**39.32**	30.24	**29.96**	21.24	**48.88**	44.45	**55.85**	45.89
		65.06	61.03	**50.82**	43.28	**38.87**	30.57	**29.65**	21.61	**48.59**	44.38	**55.08**	46.14
		65.06	61.16	**50.94**	43.29	**39.02**	30.61	**29.81**	21.72	**48.81**	44.46	**55.86**	46.23
		65.16	61.04	**50.98**	43.21	**39.04**	30.58	**29.82**	21.70	**48.83**	44.44	**55.96**	46.00

Result 3		3.3 ResNet101 vs VGG19											
		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101
BB 30k char FE	b1	30.37	**31.99**	19.24	**20.39**	11.88	**12.75**	7.65	**8.34**	28.99	**29.63**	38.22	**39.79**
	b2	29.31	**30.26**	19.20	**19.82**	12.39	**12.80**	8.32	**8.66**	30.31	**30.66**	43.84	**45.33**
	b3	29.58	**30.86**	19.42	**20.22**	12.55	**13.10**	8.45	**8.90**	30.42	**30.73**	44.30	**46.10**
	b4	29.72	**31.22**	19.53	**20.55**	12.65	**13.35**	8.53	**9.10**	30.45	**30.95**	44.33	**46.87**
	b5	29.75	**31.18**	19.54	**20.49**	12.66	**13.31**	8.54	**9.06**	30.50	**30.87**	44.49	**46.56**
BB 30k seg FE	b1	25.50	**26.66**	12.43	**12.86**	6.89	**7.09**	3.89	**4.05**	**24.07**	24.01	**28.68**	29.18
	b2	**23.57**	22.97	**12.74**	12.27	**7.49**	7.26	**4.46**	4.43	24.77	**24.99**	34.26	**37.10**
	b3	**24.30**	23.38	**13.39**	12.68	**7.95**	7.61	**4.79**	4.68	**25.30**	25.20	36.22	**37.37**
	b4	**24.67**	23.51	**13.66**	12.79	**8.17**	7.71	**4.97**	4.75	**25.51**	25.36	36.94	**37.63**
	b5	**24.72**	23.47	**13.69**	12.75	**8.19**	7.68	**4.98**	4.74	**25.53**	25.35	36.87	**37.45**
BB 8k char FE	b1	56.25	**63.01**	43.49	**49.63**	32.84	**38.40**	24.67	**29.55**	47.77	**49.56**	54.29	**59.91**
	b2	66.36	**67.05**	53.22	**54.35**	41.74	**43.20**	32.52	**34.20**	50.24	**50.89**	60.30	**63.77**
	b3	66.44	**67.67**	53.12	**54.93**	41.46	**43.59**	32.29	**34.54**	50.24	**51.21**	60.34	**64.89**
	b4	66.41	**67.62**	53.13	**54.83**	41.51	**43.50**	32.37	**34.47**	50.12	**51.04**	60.74	**64.54**
	b5	66.22	**67.49**	52.97	**54.61**	41.37	**43.20**	32.27	**34.13**	50.13	**50.89**	60.80	**64.15**
BB 8k seg FE	b1	57.35	**61.76**	39.32	**43.30**	26.51	**30.06**	18.10	**20.93**	45.02	**46.13**	44.00	**49.13**
	b2	**63.35**	62.45	46.26	**46.66**	33.40	**34.56**	24.21	**25.53**	46.25	**46.68**	49.30	**53.35**
	b3	63.00	**63.10**	46.02	**46.90**	33.41	**34.72**	24.32	**25.58**	46.45	**46.67**	49.82	**52.32**
	b4	**63.19**	63.18	46.23	**47.17**	33.56	**35.15**	24.45	**26.11**	46.52	**47.02**	49.94	**53.29**
	b5	63.28	**63.32**	46.35	**47.27**	33.70	**35.23**	24.58	**26.21**	46.60	**47.06**	49.94	**53.66**

		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101	VGG19	RN101
NB 30k char FE	b1	31.93	**32.05**	20.10	**20.42**	12.35	**12.72**	7.97	**8.31**	28.91	**29.40**	35.75	**37.31**
	b2	29.34	**29.53**	18.81	**19.39**	11.87	**12.48**	7.84	**8.40**	29.69	**30.56**	42.32	**45.08**
	b3	29.78	**30.47**	19.14	**19.99**	12.08	**12.84**	8.00	**8.66**	29.83	**30.68**	43.06	**45.35**
	b4	29.84	**30.66**	19.19	**20.19**	12.12	**13.06**	8.01	**8.87**	29.82	**30.82**	43.11	**45.78**
	b5	29.80	**30.78**	19.16	**20.26**	12.10	**13.13**	8.00	**8.92**	29.85	**30.86**	43.61	**45.79**
NB 30k seg FE	b1	**26.90**	26.53	**13.07**	13.04	7.17	**7.18**	4.02	**4.15**	24.13	**24.22**	28.79	**29.91**
	b2	**23.46**	23.22	12.29	**12.32**	6.98	**7.24**	4.07	**4.36**	24.50	**25.04**	33.74	**35.99**
	b3	23.79	**23.99**	12.59	**12.88**	7.23	**7.60**	4.25	**4.58**	24.76	**25.46**	34.37	**36.67**
	b4	**24.06**	24.03	12.73	**12.88**	7.29	**7.60**	4.28	**4.58**	24.89	**25.48**	34.45	**37.13**
	b5	**24.06**	24.05	12.72	**12.88**	7.28	**7.61**	4.26	**4.59**	24.89	**25.45**	34.41	**37.02**
NB 8k char FE	b1	60.11	**61.67**	45.61	**47.53**	33.82	**35.75**	25.04	**26.79**	47.01	**47.33**	48.77	**54.27**
	b2	65.23	**66.01**	51.27	**52.75**	39.32	**41.09**	29.96	**31.85**	48.88	**49.43**	55.85	**60.83**
	b3	65.06	**66.15**	50.82	**53.00**	38.87	**41.35**	29.65	**32.04**	48.59	**49.96**	55.08	**61.37**
	b4	65.06	**66.39**	50.94	**53.26**	39.02	**41.62**	29.81	**32.28**	48.81	**49.91**	55.86	**61.40**
	b5	65.16	**66.17**	50.98	**53.12**	39.04	**41.55**	29.82	**32.25**	48.83	**49.86**	55.96	**61.21**
NB 8k seg FE	b1	56.97	**57.91**	38.73	**39.50**	26.37	**26.76**	17.95	**17.99**	43.51	**44.54**	41.99	**44.08**
	b2	60.97	**62.78**	42.97	**44.64**	30.24	**31.43**	21.24	**22.03**	44.45	**45.72**	45.89	**49.23**
	b3	61.03	**62.67**	43.28	**44.73**	30.57	**31.58**	21.61	**22.38**	44.38	**46.01**	46.14	**50.01**
	b4	61.16	**62.66**	43.29	**44.68**	30.61	**31.46**	21.72	**22.20**	44.46	**46.03**	46.23	**50.02**
	b5	61.04	**62.68**	43.21	**44.69**	30.58	**31.50**	21.70	**22.28**	44.44	**46.12**	46.00	**50.27**

Result 4 **3.4 Feature Extraction vs Fine-Tuning**

		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		FE	FT	FE	FT	FE	FT	FE	FT	FE	FT	FE	FT
BB 30k char RN	b1	**31.99**	30.91	**20.39**	19.74	**12.75**	12.22	**8.34**	7.92	**29.63**	29.43	**39.79**	37.99
	b2	30.26	**31.16**	19.82	**20.67**	12.80	**13.39**	8.66	**9.06**	30.66	**31.09**	45.33	**46.52**
	b3	30.86	**31.53**	**20.22**	21.02	13.10	**13.75**	8.90	**9.39**	30.73	**31.16**	46.10	**46.11**
	b4	31.22	**31.51**	20.55	**21.06**	13.35	**13.86**	9.10	**9.53**	30.95	**31.25**	**46.87**	46.20
	b5	31.18	**31.53**	20.49	**21.11**	13.31	**13.92**	9.06	**9.58**	30.87	**31.26**	**46.56**	46.23
BB 30k seg RN	b1	**26.66**	26.62	12.86	**13.32**	7.09	**7.45**	4.05	**4.33**	24.01	**25.35**	29.18	**32.73**
	b2	22.97	**23.54**	12.27	**12.80**	7.26	**7.67**	4.43	**4.76**	24.99	**25.77**	37.10	**37.15**
	b3	23.38	**24.34**	12.68	**13.42**	7.61	**8.10**	4.68	**4.99**	25.20	**26.28**	37.37	**38.17**
	b4	23.51	**24.28**	12.79	**13.46**	7.71	**8.15**	4.75	**5.07**	25.36	**26.26**	37.63	**38.10**
	b5	23.47	**24.24**	12.75	**13.44**	7.68	**8.16**	4.74	**5.06**	25.35	**26.26**	37.45	**38.25**
BB 8k char RN	b1	**63.01**	59.05	**49.63**	45.72	**38.40**	34.50	**29.55**	25.90	**49.56**	48.34	**59.91**	53.59
	b2	67.05	**68.23**	54.35	**55.22**	43.20	**43.75**	34.20	**34.61**	50.89	**51.03**	**63.77**	62.17
	b3	67.67	**68.22**	54.93	**55.41**	43.59	**44.10**	34.54	**34.99**	51.21	**51.43**	**64.89**	63.75
	b4	67.62	**68.19**	54.83	**55.29**	43.50	**43.95**	34.47	**34.88**	51.04	**51.33**	**64.54**	63.27
	b5	67.49	**68.21**	54.61	**55.38**	43.20	**44.12**	34.13	**35.08**	50.89	**51.39**	**64.15**	63.54
BB 8k seg RN	b1	61.76	**65.34**	43.30	**46.03**	30.06	**32.10**	20.93	**22.57**	46.13	**47.15**	**49.13**	49.10
	b2	62.45	**64.16**	46.66	**48.05**	34.56	**35.52**	25.53	**26.23**	46.68	**47.64**	**53.35**	51.63
	b3	63.10	**65.00**	46.90	**48.80**	34.72	**36.10**	25.58	**26.69**	46.67	**47.83**	**52.32**	51.74
	b4	63.18	**65.00**	47.17	**48.71**	35.15	**36.02**	26.11	**26.56**	47.02	**48.06**	**53.29**	52.38
	b5	63.32	**64.98**	47.27	**48.63**	35.23	**35.91**	26.21	**26.47**	47.06	**47.94**	**53.66**	52.06

		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		FE	FT	FE	FT	FE	FT	FE	FT	FE	FT	FE	FT
NB 30k char RN	b1	32.05	**33.18**	20.42	**21.36**	12.72	**13.36**	8.31	**8.70**	29.40	**29.63**	37.31	**38.97**
	b2	**29.53**	29.19	**19.39**	19.12	**12.48**	12.22	**8.40**	8.15	**30.56**	29.96	**45.08**	43.19
	b3	**30.47**	29.87	**19.99**	19.54	**12.84**	12.49	**8.66**	8.31	**30.68**	30.06	**45.35**	42.87
	b4	**30.66**	29.89	**20.19**	19.55	**13.06**	12.52	**8.87**	8.34	**30.82**	30.10	**45.78**	42.62
	b5	**30.78**	29.91	**20.26**	19.57	**13.13**	12.52	**8.92**	8.34	**30.86**	30.15	**45.79**	42.71
NB 30k seg RN	b1	26.53	**26.98**	13.04	**13.40**	7.18	**7.35**	**4.15**	4.13	24.22	**24.66**	29.91	**30.01**
	b2	**23.22**	23.11	**12.32**	11.98	**7.24**	6.91	**4.36**	4.13	**25.04**	24.51	**35.99**	34.03
	b3	**23.99**	23.64	**12.88**	12.40	**7.60**	7.22	**4.58**	4.31	**25.46**	24.80	**36.67**	34.43
	b4	**24.03**	23.54	**12.88**	12.36	**7.60**	7.22	**4.58**	4.31	**25.48**	24.78	**37.13**	34.29
	b5	**24.05**	23.54	**12.88**	12.37	**7.61**	7.24	**4.59**	4.34	**25.45**	24.80	**37.02**	34.59
NB 8k char RN	b1	**61.67**	59.06	**47.53**	44.92	**35.75**	33.27	**26.79**	24.46	**47.33**	46.35	**54.27**	46.78
	b2	**66.01**	64.49	**52.75**	50.43	**41.09**	38.40	**31.85**	29.11	**49.43**	47.64	**60.83**	51.98
	b3	**66.15**	64.60	**53.00**	50.60	**41.35**	38.57	**32.04**	29.27	**49.96**	47.97	**61.37**	52.54
	b4	**66.39**	64.76	**53.26**	50.75	**41.62**	38.74	**32.28**	29.46	**49.91**	48.09	**61.40**	52.74
	b5	**66.17**	64.71	**53.12**	50.67	**41.55**	38.67	**32.25**	29.38	**49.86**	47.99	**61.21**	52.41
NB 8k seg RN	b1	57.91	**59.98**	39.50	**40.55**	26.76	**27.79**	17.99	**19.39**	**44.54**	44.32	**44.08**	41.74
	b2	**62.78**	62.08	**44.64**	44.14	31.43	**31.49**	22.03	**22.84**	**45.72**	45.25	**49.23**	45.72
	b3	**62.67**	62.42	**44.73**	44.48	31.58	**31.71**	22.38	**23.01**	**46.01**	45.71	**50.01**	46.75
	b4	**62.66**	62.49	**44.68**	44.49	31.46	**31.71**	22.20	**23.00**	**46.03**	45.85	**50.02**	47.06
	b5	**62.68**	62.43	**44.69**	44.49	31.50	**31.76**	22.28	**23.06**	**46.12**	45.90	**50.27**	47.18

Result 5		3.5 Checkpoints with Best BLEU Score vs Continue Training											
		B-1		B-2		B-3		B-4		ROUGE_L		CIDEr	
		BB	NB	BB	NB	BB	NB	BB	NB	BB	NB	BB	NB
30k char FE RN	b1	31.99	**32.05**	20.39	20.42	**12.75**	12.72	**8.34**	8.31	**29.63**	29.40	**39.79**	37.31
	b2	**30.26**	29.53	19.82	19.39	**12.80**	12.48	**8.66**	8.40	**30.66**	30.56	**45.33**	45.08
	b3	**30.86**	30.47	20.22	19.99	**13.10**	12.84	**8.90**	8.66	**30.73**	30.68	**46.10**	45.35
	b4	**31.22**	30.66	20.55	20.19	**13.35**	13.06	**9.10**	8.87	**30.95**	30.82	**46.87**	45.78
	b5	**31.18**	30.78	20.49	20.26	**13.31**	13.13	**9.06**	8.92	**30.87**	30.86	**46.56**	45.79
30k char FT RN	b1	30.91	**33.18**	19.74	21.36	12.22	**13.36**	7.92	**8.70**	29.43	**29.63**	37.99	**38.97**
	b2	**31.16**	29.19	**20.67**	19.12	**13.39**	12.22	**9.06**	8.15	**31.09**	29.96	**46.52**	43.19
	b3	**31.53**	29.87	**21.02**	19.54	**13.75**	12.49	**9.39**	8.31	**31.16**	30.06	**46.11**	42.87
	b4	**31.51**	29.89	**21.06**	19.55	**13.86**	12.52	**9.53**	8.34	**31.25**	30.10	**46.20**	42.62
	b5	**31.53**	29.91	**21.11**	19.57	**13.92**	12.52	**9.58**	8.34	**31.26**	30.15	**46.23**	42.71
30k seg FE RN	b1	**26.66**	26.53	12.86	**13.04**	7.09	**7.18**	4.05	**4.15**	24.01	**24.22**	29.18	**29.91**
	b2	22.97	**23.22**	12.27	**12.32**	**7.26**	7.24	**4.43**	4.36	24.99	**25.04**	**37.10**	35.99
	b3	23.38	**23.99**	12.68	**12.88**	**7.61**	7.60	**4.68**	4.58	25.20	**25.46**	**37.37**	36.67
	b4	23.51	**24.03**	12.79	**12.88**	**7.71**	7.60	**4.75**	4.58	25.36	**25.48**	**37.63**	37.13
	b5	23.47	**24.05**	12.75	**12.88**	**7.68**	7.61	**4.74**	4.59	25.35	**25.45**	**37.45**	37.02
30k seg FT RN	b1	26.62	**26.98**	13.32	**13.40**	**7.45**	7.35	**4.33**	4.13	**25.35**	24.66	**32.73**	30.01
	b2	**23.54**	23.11	**12.80**	11.98	**7.67**	6.91	**4.76**	4.13	**25.77**	24.51	**37.15**	34.03
	b3	**24.34**	23.64	**13.42**	12.40	**8.10**	7.22	**4.99**	4.31	**26.28**	24.80	**38.17**	34.43
	b4	**24.28**	23.54	**13.46**	12.36	**8.15**	7.22	**5.07**	4.31	**26.26**	24.78	**38.10**	34.29
	b5	**24.24**	23.54	**13.44**	12.37	**8.16**	7.24	**5.06**	4.34	**26.26**	24.80	**38.25**	34.59
8k char FE RN	b1	**63.01**	61.67	**49.63**	47.53	**38.40**	35.75	**29.55**	26.79	**49.56**	47.33	**59.91**	54.27
	b2	**67.05**	66.01	**54.35**	52.75	**43.20**	41.09	**34.20**	31.85	**50.89**	49.43	**63.77**	60.83
	b3	**67.67**	66.15	**54.93**	53.00	**43.59**	41.35	**34.54**	32.04	**51.21**	49.96	**64.89**	61.37
	b4	**67.62**	66.39	**54.83**	53.26	**43.50**	41.62	**34.47**	32.28	**51.04**	49.91	**64.54**	61.40
	b5	**67.49**	66.17	**54.61**	53.12	**43.20**	41.55	**34.13**	32.25	**50.89**	49.86	**64.15**	61.21
8k char FT RN	b1	59.05	**59.06**	**45.72**	44.92	**34.50**	33.27	**25.90**	24.46	**48.34**	46.35	**53.59**	46.78
	b2	**68.23**	64.49	**55.22**	50.43	**43.75**	38.40	**34.61**	29.11	**51.03**	47.64	**62.17**	51.98
	b3	**68.22**	64.60	**55.41**	50.60	**44.10**	38.57	**34.99**	29.27	**51.43**	47.97	**63.75**	52.54
	b4	**68.19**	64.76	**55.29**	50.75	**43.95**	38.74	**34.88**	29.46	**51.33**	48.09	**63.27**	52.74
	b5	**68.21**	64.71	**55.38**	50.67	**44.12**	38.67	**35.08**	29.38	**51.39**	47.99	**63.54**	52.41
8k seg FE RN	b1	**61.76**	57.91	**43.30**	39.50	**30.06**	26.76	**20.93**	17.99	**46.13**	44.54	**49.13**	44.08
	b2	62.45	**62.78**	**46.66**	44.64	**34.56**	31.43	**25.53**	22.03	**46.68**	45.72	**53.35**	49.23
	b3	**63.10**	62.67	**46.90**	44.73	**34.72**	31.58	**25.58**	22.38	**46.67**	46.01	**52.32**	50.01
	b4	**63.18**	62.66	**47.17**	44.68	**35.15**	31.46	**26.11**	22.20	**47.02**	46.03	**53.29**	50.02
	b5	**63.32**	62.68	**47.27**	44.69	**35.23**	31.50	**26.21**	22.28	**47.06**	46.12	**53.66**	50.27
8k seg FT RN	b1	**65.34**	59.98	**46.03**	40.55	**32.10**	27.79	**22.57**	19.39	**47.15**	44.32	**49.10**	41.74
	b2	**64.16**	62.08	**48.05**	44.14	**35.52**	31.49	**26.23**	22.84	**47.64**	45.25	**51.63**	45.72
	b3	**65.00**	62.42	**48.80**	44.48	**36.10**	31.71	**26.69**	23.01	**47.83**	45.71	**51.74**	46.75
	b4	**65.00**	62.49	**48.71**	44.49	**36.02**	31.71	**26.56**	23.00	**48.06**	45.85	**52.38**	47.06
	b5	**64.98**	62.43	**48.63**	44.49	**35.91**	31.76	**26.47**	23.06	**47.94**	45.90	**52.06**	47.18
30k char FE VG	b1	30.37	**31.93**	19.24	**20.10**	11.88	**12.35**	7.65	**7.97**	**28.99**	28.91	**38.22**	35.75
	b2	29.31	**29.34**	**19.20**	18.81	**12.39**	11.87	**8.32**	7.84	**30.31**	29.69	**43.84**	42.32
	b3	29.58	**29.78**	**19.42**	19.14	**12.55**	12.08	**8.45**	8.00	**30.42**	29.83	**44.30**	43.06
	b4	29.72	**29.84**	**19.53**	19.19	**12.65**	12.12	**8.53**	8.01	**30.45**	29.82	**44.33**	43.11
	b5	29.75	**29.80**	**19.54**	19.16	**12.66**	12.10	**8.54**	8.00	**30.50**	29.85	**44.49**	43.61
30k seg FE VG	b1	25.50	**26.90**	**12.43**	13.07	6.89	**7.17**	3.89	**4.02**	24.07	**24.13**	28.68	**28.79**
	b2	**23.57**	23.46	**12.74**	12.29	**7.49**	6.98	**4.46**	4.07	**24.77**	24.50	**34.26**	33.74
	b3	**24.30**	23.79	**13.39**	12.59	**7.95**	7.23	**4.79**	4.25	**25.30**	24.76	**36.22**	34.37
	b4	**24.67**	24.06	**13.66**	12.73	**8.17**	7.29	**4.97**	4.28	**25.51**	24.89	**36.94**	34.45
	b5	**24.72**	24.06	**13.69**	12.72	**8.19**	7.28	**4.98**	4.26	**25.53**	24.89	**36.87**	34.41
8k char FE VG	b1	56.25	**60.11**	43.49	**45.61**	32.84	**33.82**	24.67	**25.04**	**47.77**	47.01	**54.29**	48.77
	b2	**66.36**	65.23	**53.22**	51.27	**41.74**	39.32	**32.52**	29.96	**50.24**	48.88	**60.30**	55.85
	b3	**66.44**	65.06	**53.12**	50.82	**41.46**	38.87	**32.29**	29.65	**50.24**	48.59	**60.34**	55.08
	b4	**66.41**	65.06	**53.13**	50.94	**41.51**	39.02	**32.37**	29.81	**50.12**	48.81	**60.74**	55.86
	b5	**66.22**	65.16	**52.97**	50.98	**41.37**	39.04	**32.27**	29.82	**50.13**	48.83	**60.80**	55.96
8k seg FE VG	b1	**57.35**	56.97	**39.32**	38.73	**26.51**	26.37	**18.10**	17.95	**45.02**	43.51	**44.00**	41.99
	b2	**63.35**	60.97	**46.26**	42.97	**33.40**	30.24	**24.21**	21.24	**46.25**	44.45	**49.30**	45.89
	b3	**63.00**	61.03	**46.02**	43.28	**33.41**	30.57	**24.32**	21.61	**46.45**	44.38	**49.82**	46.14
	b4	**63.19**	61.16	**46.23**	43.29	**33.56**	30.61	**24.45**	21.72	**46.52**	44.46	**49.94**	46.23
	b5	**63.28**	61.04	**46.35**	43.21	**33.70**	30.58	**24.58**	21.70	**46.60**	44.44	**49.94**	46.00

Captioning Examples 1

BB	30k	seg	FE	RN	'一个', '穿', '着', '蓝色', '衬衫', '的', '年轻', '男孩', '在', '看', '他', '的', '手机'
					A young boy in a blue shirt is looking at his mobile phone
BB	30k	char	FE	RN	'一', '个', '年', '轻', '的', '女', '人', '看', '着', '她', '的', '电', '话'
					A young woman is looking at her telephone
BB	8k	seg	FE	RN	**'一个', '穿', '着', '蓝色', '衬衫', '的', '男人', '在', '看', '着', '他', '的', '手机'**
					A man in a blue shirt is looking at his mobile phone
BB	8k	char	FE	RN	**'一', '个', '小', '女', '孩', '看', '了', '看', '相', '机'**
					A young girl looks at camera
BB	30k	seg	FT	RN	**'一个', '戴', '着', '眼镜', '的', '人', '在', '看', '一', '台', '笔记本', '电脑'**
					A person in glasses is looking at a notebook
BB	30k	char	FT	RN	**'一', '个', '穿', '着', '蓝', '色', '衬', '衫', '的', '年', '轻', '女', '孩', '正', '在', '看', '书'**
					A young girl in blue shirt is reading book
BB	8k	seg	FT	RN	**'一个', '穿', '着', '蓝色', '衬衫', '的', '年轻', '女子', '在', '她', '的', '头', '上', '看', '着', '相机'**
					A young girl in a blue shirt is on her head looking at camera
BB	8k	char	FT	RN	'一', '个', '穿', '着', '蓝', '色', '衬', '衫', '的', '年', '轻', '女', '孩', '坐', '在', '一', '个', '白', '色', '的', '窗', '台', '上'
					A young girl in a white shirt is sitting on a white windowsill
BB	30k	seg	FE	VG	'一个', '穿', '着', '蓝色', '衬衫', '的', '女人', '在', '看', '一', '本', '书'
					A woman in a blue shirt is reading a book
BB	30k	char	FE	VG	'一', '个', '年', '轻', '的', '男', '孩', '在', '一', '个', '黑', '色', '的', '衬', '衫', '和', '一', '个', '黑', '色', '的', '衬', '衫', '在', '一', '个', '电', '脑', '屏', '幕', '上', '看'
					A young boy in a blue shirt in a blue shirt is on a computer screen looking
BB	8k	seg	FE	VG	**'一个', '小', '女孩', '坐', '在', '一', '张', '桌子', '上'**
					A young girl is sitting on a table
BB	8k	char	FE	VG	**'一', '个', '年', '轻', '的', '女', '孩', '坐', '在', '一', '个', '电', '脑', '上'**
					A young girl is sitting on a computer

Captioning Examples 2

BB	30k	seg	FE	RN	'一个', '人', '在', '一个', '黄色', '的', '冲浪', '板'
					A person at a yellow surfing board
BB	30k	char	FE	RN	'一', '个', '男', '人', '骑', '着', '一', '个', '黄', '色', '的', '冲', '浪', '板', '冲', '浪'
					A man riding a yellow surfing board surfs
BB	8k	seg	FE	RN	'一个', '穿', '着', '黄色', '衬衫', '的', '男孩', '在', '冲浪', '板'
					A boy in a yellow shirt is surfing board
BB	8k	char	FE	RN	'冲', '浪', '者', '骑', '波'
					Surfer riding waves
BB	30k	seg	FT	RN	'在', '一个', '绿色', '的', '潜水', '衣', '一个', '人', '在', '冲浪', '板'
					At a green diving clothes a person at surfing board
BB	30k	char	FT	RN	'一', '个', '男', '人', '在', '一', '个', '绿', '色', '的', '冲', '浪', '板', '冲', '浪'
					A man at a green surfing board surfs
BB	8k	seg	FT	RN	'在', '一个', '黄色', '的', '潜水', '服', '的', '人'
					person at a yellow diving clothes
BB	8k	char	FT	RN	**'一', '个', '小', '男', '孩', '在', '海', '洋', '中', '冲', '浪'**
					A little boy is surfing in the ocean
BB	30k	seg	FE	VG	**'一个', '人', '在', '海洋', '中', '冲浪'**
					A person is surfing in the ocean
BB	30k	char	FE	VG	**'一', '个', '年', '轻', '的', '冲', '浪', '者', '骑', '波'**
					A young surfer rides waves
BB	8k	seg	FE	VG	**'在', '一个', '黄色', '的', '冲浪', '板', '冲浪', '的', '人'**
					person at a yellow surfing board
BB	8k	char	FE	VG	'一', '个', '男', '孩', '在', '冲', '浪', '板', '上', '冲', '浪'
					a boy surfs on a surfing board

Captioning Examples 3

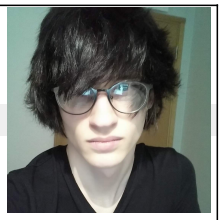
BB	30k	seg	FE	RN	'一', '只', '黄色', '的', '花', '在', '一', '片', '黄色', '的', '花', '上'
					A yellow flower at a field of yellow flowers
BB	30k	char	FE	RN	'一', '个', '穿', '着', '黄', '色', '衣', '服', '的', '女', '孩', '正', '站', '在', '一', '个', '黄', '色', '的', '花', '旁', '边'
					A girl in yellow clothes is standing beside a yellow flower
BB	8k	seg	FE	RN	'一', '只', '黄色', '的', '花', '在', '一个', '黄色', '的', '花'
					A yellow flower at a yellow flower
BB	8k	char	FE	RN	'一', '只', '白', '色', '的', '狗', '在', '一', '个', '黄', '色', '的', '花', '朵', '上'
					A white dog on a yellow flower
BB	30k	seg	FT	RN	'一个', '人', '坐', '在', '一', '把', '黄色', '的', '椅子', '上'
					A person sitting on a yellow chair
BB	30k	char	FT	RN	'一', '群', '人', '在', '一', '个', '黄', '色', '的', '郁', '金', '香'
					A group of people at a yellow tulip
BB	8k	seg	FT	RN	'一个', '人', '在', '一个', '黄色', '的', '帐篷', '里', '玩耍'
					A person is playing in a yellow tent
BB	8k	char	FT	RN	'一', '个', '穿', '着', '黄', '色', '衬', '衫', '的', '人', '在', '一', '个', '黄', '色', '的', '花', '园', '里'
					A person in yellow shirt in a yellow garden
BB	30k	seg	FE	VG	'一', '群', '人', '坐', '在', '一个', '黄色', '的', '<unk>'
					A group of people sitting at a yellow <unk>
BB	30k	char	FE	VG	'一', '个', '穿', '着', '黄', '色', '衬', '衫', '的', '人', '在', '一', '个', '黄', '色', '的', '田', '野', '里'
					A person in yellow shirt in a yellow field
BB	8k	seg	FE	VG	'一', '只', '黄色', '的', '狗', '在', '一个', '黄色', '的', '黄色', '和', '黄色', '的', '<unk>'
					A yellow dog at a yellow and yellow <unk>
BB	8k	char	FE	VG	'一', '个', '黄', '色', '的', '毛', '茸', '茸', '的', '黄', '色', '的', '狗', '在', '草', '地', '上', '玩'
					A yellow furry yellow dog playing on the grass

Captioning Examples 4

BB	30k	seg	FE	RN	'一个', '小', '女孩', '骑', '着', '红色', '三轮车'
					a young girl is riding a red tricycle
BB	30k	char	FE	RN	'一', '个', '小', '女', '孩', '骑', '着', '一', '辆', '红', '色', '的', '自', '行', '车'
					a young girl is riding a red bicycle
BB	8k	seg	FE	RN	'一个', '小', '女孩', '骑', '着', '一', '辆', '红色', '的', '三轮车'
					a young girl is riding a red tricycle
BB	8k	char	FE	RN	'一', '个', '小', '女', '孩', '骑', '着', '三', '轮', '车'
					a young girl is riding tricycle
BB	30k	seg	FT	RN	'一个', '小', '男孩', '骑', '着', '一', '辆', '摩托车'
					a young boy is riding a motorcycle
BB	30k	char	FT	RN	'一', '个', '小', '女', '孩', '骑', '着', '一', '辆', '摩', '托', '车'
					a young girl is riding a motorcycle
BB	8k	seg	FT	RN	'一个', '小', '女孩', '骑', '着', '三轮车'
					a young girl is riding tricycle
BB	8k	char	FT	RN	'一', '个', '小', '女', '孩', '骑', '着', '独', '轮', '车'
					a young is riding a unicycle
BB	30k	seg	FE	VG	'一个', '小', '男孩', '骑', '着', '一', '辆', '红色', '的', '自行车'
					a young boy is riding a red bibycle
BB	30k	char	FE	VG	'一', '个', '小', '男', '孩', '骑', '着', '一', '辆', '红', '色', '的', '自', '行', '车'
					a young boy is riding a red bicycle
BB	8k	seg	FE	VG	'一个', '小', '女孩', '骑', '着', '一', '辆', '红色', '的', '玩具'
					a young girl is riding a red toy
BB	8k	char	FE	VG	'一', '个', '穿', '着', '红', '色', '衬', '衫', '的', '小', '男', '孩', '骑', '着', '一', '辆', '红', '色', '的', '自', '行', '车'
					a boy in red shirt is riding a red bicycle

Captioning Examples 5

BB	30k	seg	FE	RN	'一个', '戴', '着', '墨镜', '的', '男人'
					a man in sunglasses
BB	30k	char	FE	RN	'一', '个', '戴', '着', '墨', '镜', '的', '男', '孩', '戴', '着', '墨', '镜'
					a boy in sunglasses is wearing a pair of sunglasses
BB	8k	seg	FE	RN	'一个', '戴', '着', '墨镜', '的', '女人'
					a woman in sunglasses
BB	8k	char	FE	RN	'一', '个', '戴', '着', '墨', '镜', '的', '男', '人'
					a man in sunglasses
BB	30k	seg	FT	RN	'一个', '戴', '着', '眼镜', '的', '年轻', '女孩'
					a young girl in sunglasses
BB	30k	char	FT	RN	'一', '个', '戴', '着', '墨', '镜', '的', '男', '孩'
					a boy in sunglasses
BB	8k	seg	FT	RN	'一个', '戴', '着', '墨镜', '的', '女人'
					a woman in sunglasses
BB	8k	char	FT	RN	'一', '个', '戴', '着', '墨', '镜', '的', '女', '人'
					a woman in sunglasses
BB	30k	seg	FE	VG	'一', '个', '戴', '着', '墨', '镜', '的', '女', '人',
					a woman in sunglasses
BB	30k	char	FE	VG	'一', '个', '戴', '着', '墨', '镜', '的', '男', '人'
					a man in sunglasses
BB	8k	seg	FE	VG	'戴', '眼镜', '的', '女人'
					woman in sunglasses
BB	8k	char	FE	VG	'一', '个', '戴', '着', '墨', '镜', '的', '男', '人', '和', '一', '个', '女', '人'
					a man in sunglasses and a woman



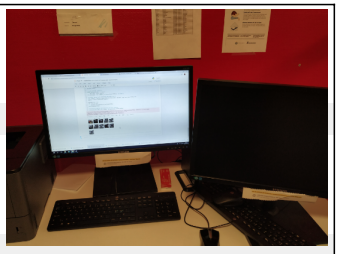
Captioning Examples 6

BB	30k	seg	FE	RN	'一个', '小', '女孩', '坐', '在', '一', '张', '桌子', '上', '用', '一', '把', '勺子', '在', '碗', '里', '搅拌'
					a little girl sitting on a table, using a spone stiring in a bowl
BB	30k	char	FE	RN	'一', '个', '穿', '着', '粉', '红', '色', '衣', '服', '的', '婴', '儿', '坐', '在', '一', '个', '碗', '里'
					a baby in pink clothes sitting in a bowl
BB	8k	seg	FE	RN	'一个', '婴儿', '坐', '在', '一', '张', '桌子', '上'
					a baby sitting on a table
BB	8k	char	FE	RN	'一', '个', '微', '笑', '的', '婴', '儿', '坐', '在', '一', '个', '<unk>', '<unk>', '的', '桌', '子', '上'
					a smiling baby sitting on a <unk> <unk> table
BB	30k	seg	FT	RN	'一个', '穿', '着', '睡衣', '的', '小', '女孩', '坐', '在', '一个', '粉红色', '的', '碗', '里'
					a little gir in sleeping clothes sitting in a pink bowl
BB	30k	char	FT	RN	'一', '个', '小', '女', '孩', '坐', '在', '一', '个', '粉', '红', '色', '的', '椅', '子', '上'
					a little girl sitting on a pink chair
BB	8k	seg	FT	RN	'一个', '小', '女孩', '坐', '在', '一个', '红色', '的', '椅子', '上'
					a little girl on a pink chair
BB	8k	char	FT	RN	'一', '个', '小', '男', '孩', '坐', '在', '一', '个', '红', '色', '的', '垫', '子', '上'
					a little boy sitting on a pink mat
BB	30k	seg	FE	VG	'一个', '小', '女孩', '坐', '在', '一', '张', '桌子', '上'
					a little girl sitting on a table
BB	30k	char	FE	VG	'一', '个', '年', '轻', '的', '亚', '洲', '女', '孩', '坐', '在', '一', '个', '白', '色', '的', '碗', '里'
					a young Asian girl sitting in a white bowl
BB	8k	seg	FE	VG	'一个', '穿', '着', '粉红色', '连衣裙', '的', '小', '女孩', '坐', '在', '一个', '蓝色', '的', '幻灯片'
					a little girl in pink dress sitting a blue slideshow
BB	8k	char	FE	VG	'一', '个', '穿', '着', '粉', '红', '色', '的', '小', '女', '孩', '坐', '在', '一', '个', '红', '色', '的', '玩', '具'
					a little girl in pink sitting a red toy

Captioning Examples 7

BB	30k	seg	FE	RN	'一', '只', '棕色', '的', '狗', '正', '站', '在', '一个', '装满', '水果', '的', '篮子', '里'
					a brown dog is standing in a basket full of fruits
BB	30k	char	FE	RN	'一', '个', '女', '人', '坐', '在', '一', '个', '水', '果', '摊', '旁', '边'
					a woman is beside a fruit stall
BB	8k	seg	FE	RN	'一个', '小', '男孩', '坐', '在', '一', '张', '桌子', '上'
					a little is sitting on a table
BB	8k	char	FE	RN	'一', '个', '特', '写', '镜', '头', '的', '男', '孩'
					a close-up of boy
BB	30k	seg	FT	RN	'一个', '人', '在', '一个', '大', '的', '南瓜', '上', '休息'
					a person is resting on a big pumpkin
BB	30k	char	FT	RN	'一', '个', '人', '在', '一', '个', '黄', '色', '的', '幻', '灯', '片', '上'
					a person on a yellow slideshow
BB	8k	seg	FT	RN	'一个', '小', '男孩', '在', '一个', '黄色', '的', '管子', '里', '玩'
					a little boy is playing in a yellow pipeline
BB	8k	char	FT	RN	'一', '个', '人', '坐', '在', '一', '个', '黄', '色', '的', '球'
					a person sitting a yellow ball
BB	30k	seg	FE	VG	'一个', '穿', '着', '黄色', '衬衫', '的', '人', '坐', '在', '一个', '篮子', '里'
					a person in yellow shirt sitting in a basket
BB	30k	char	FE	VG	'一', '个', '年', '轻', '的', '女', '孩', '坐', '在', '一', '个', '篮', '子', '里'
					a young girl is sitting in a basket
BB	8k	seg	FE	VG	'一个', '小', '女孩', '在', '一个', '黄色', '的', '球'
					a little girl at a yellow ball
BB	8k	char	FE	VG	'一', '个', '黄', '色', '的', '孩', '子', '在', '一', '个', '黄', '色', '的', '玩', '具'
					a yellow child at a yellow toy

Captioning Examples 8



BB	30k	seg	FE	RN	'一个', '人', '在', '电脑', '上', '工作'
					a person is working on computer
BB	30k	char	FE	RN	'一', '个', '人', '在', '电', '脑', '上', '工', '作'
					a person is working on computer
BB	8k	seg	FE	RN	'一个', '人', '坐', '在', '电脑', '上', '看', '书'
					a person is sitting on a computer reading book
BB	8k	char	FE	RN	'一', '个', '人', '坐', '在', '电', '脑', '上', '看', '书'
					a person is sitting on a computer reading book
BB	30k	seg	FT	RN	'一个', '人', '在', '电脑', '上', '工作'
					a person is working on computer
BB	30k	char	FT	RN	'一', '个', '人', '在', '一', '个', '电', '脑', '屏', '幕', '上', '看', '了', '一', '张', '照', '片'
					a person looking at a photo on a computer screen
BB	8k	seg	FT	RN	'一个', '人', '坐', '在', '一个', '<unk>', '的', '窗口'
					a person sitting at a <unk> window
BB	8k	char	FT	RN	'在', '一', '个', '白', '色', '的', '建', '筑', '物', '前', '面', '的', '人'
					a person in front of a white building
BB	30k	seg	FE	VG	'一个', '人', '在', '电脑', '上', '工作'
					a person is working on computer
BB	30k	char	FE	VG	'一', '个', '人', '在', '电', '脑', '上', '工', '作'
					a person is working on computer
BB	8k	seg	FE	VG	'一个', '人', '坐', '在', '一', '张', '桌子', '上'
					a person is sitting on a table
BB	8k	char	FE	VG	'一', '个', '人', '坐', '在', '电', '脑', '上'
					a person is working on computer

Captioning Examples 9

BB	30k	seg	FE	RN	'一个', '人', '在', '码头', '上', '散步'
					a person walking on a shipside
BB	30k	char	FE	RN	'一', '个', '人', '站', '在', '一', '座', '桥', '上', '一', '座', '桥'
					a person stands on a bridge a bridge
BB	8k	seg	FE	RN	'两', '只', '人', '站', '在', '一', '座', '桥', '上'
					two piece of persons standing on a bridge
BB	8k	char	FE	RN	'一', '个', '人', '站', '在', '一', '座', '桥', '上', '俯', '瞰', '着', '水'
					a person standing on a bridge looking down at the water
BB	30k	seg	FT	RN	'一个', '人', '站', '在', '桥', '上', '看', '水'
					a person standing on the bridge looking at water
BB	30k	char	FT	RN	'一', '个', '人', '在', '一', '座', '桥', '上', '看', '一', '座', '桥'
					a person on a bridge looking at a bridge
BB	8k	seg	FT	RN	'一个', '人', '站', '在', '一', '条', '长凳', '上'
					a person standing on a long bench
BB	8k	char	FT	RN	'一', '个', '人', '站', '在', '一', '个', '木', '制', '的', '长', '椅', '上'
					a person standing on a wooden bench
BB	30k	seg	FE	VG	'一个', '人', '站', '在', '一', '座', '桥', '上'
					a person standing on a bridge
BB	30k	char	FE	VG	'一', '个', '男', '人', '和', '一', '个', '女', '人', '站', '在', '一', '座', '桥', '上'
					a man and a woman standing on a bridge
BB	8k	seg	FE	VG	'两', '只', '人', '站', '在', '一', '条', '桥', '上', '走', '在', '水', '里'
					two piece of persons standing on a bridge walking in the water
BB	8k	char	FE	VG	'一', '个', '男', '人', '站', '在', '一', '个', '码', '头', '上', '看', '水'
					a man standing on a shipside looking at water

Captioning Examples 10

BB	30k	seg	FE	RN	'一', '只', '白色', '的', '狗', '嘴里', '叼', '着', '一个', '玩具'
					a white dog has a toy in its mouth
BB	30k	char	FE	RN	'一', '只', '白', '色', '的', '狗', '在', '它', '的', '嘴', '里', '叼', '着', '一', '个', '紫', '色', '的', '飞', '盘'
					a white dog has a purple frisbee in its mouth
BB	8k	seg	FE	RN	'一', '只', '白色', '的', '狗', '带', '着', '紫色', '的', '飞盘'
					a white dog holds a purple frisbee
BB	8k	char	FE	RN	'一', '只', '白', '色', '的', '狗', '叼', '着', '一', '个', '紫', '色', '的', '飞', '盘'
					a white dog has a purple frisbee in its mouth
BB	30k	seg	FT	RN	'一', '只', '白色', '的', '狗', '正在', '玩', '一个', '紫色', '的', '<unk>'
					a white a dog is playing a purple <unk>
BB	30k	char	FT	RN	'一', '只', '白', '色', '的', '狗', '拿', '着', '一', '个', '紫', '色', '的', '飞', '盘'
					a white dog holds a purple frisbee
BB	8k	seg	FT	RN	'一', '只', '白色', '的', '狗', '在', '草地', '上', '玩耍'
					a white dog is playing on the grass
BB	8k	char	FT	RN	'一', '只', '白', '色', '的', '狗', '带', '着', '一', '个', '紫', '色', '的', '圆', '盘'
					a white dog holds a purple round plate
BB	30k	seg	FE	VG	'一', '只', '白色', '的', '狗', '在', '草地', '上', '玩耍'
					a white dog plays on the grass
BB	30k	char	FE	VG	'一', '只', '白', '色', '的', '狗', '在', '草', '地', '上', '奔', '跑'
					a white dog is running on the grass
BB	8k	seg	FE	VG	'一', '只', '白色', '的', '狗', '在', '一个', '蓝色', '的', '隧道', '里'
					a white dog is in a blue tunnel
BB	8k	char	FE	VG	'一', '只', '白', '色', '的', '狗', '带', '着', '一', '个', '红', '色', '的', '飞', '盘'
					a white dog holds a red frisbee