Image Captioning in Indian Languages

BTP PHASE-I

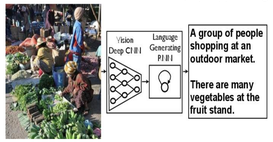
Rishabh Sharma(180102058)
Under the Guidance of Prof. Prithwijit Guha
Department of EEE, IIT Guwahati

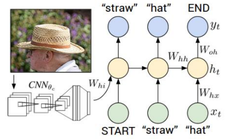
Abstract

Captioning an image with proper descriptions has become a fascinating problem. It connects computer vision and natural language processing. We explore the show and tell image caption generation model along with translater APIs which attempt to generate captions for any language.

Introduction

we propose to follow this formula, by supplanting the encoder recurrent neural network with a profound convolution





Review of Prior Works

A lot of work has been done on the picture inscription problem.

- The primary critical work in explaining picture inscribing undertakings was finished by Ali Farhadi [8] .**Mapping of Images to meaning triplet of object action scene :**

- O.Vinyals and team [1], in the work , introduced a novel approach of using (CNN) and (RNN) for image captioning tasks.

$$x_{-1} = \text{CNN}(I)$$
$$x_t = W_e S_t, \quad t \in \{0 \dots N-1\}$$
$$p_{t+1} = \text{LSTM}(x_t), \quad t \in \{0 \dots N-1\}$$

$$L(I,S) = -\sum_{t=1}^{N} \log p_t(S_t) .$$

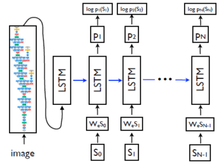
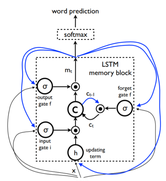


Fig 3: Model architecture of Show and Tell



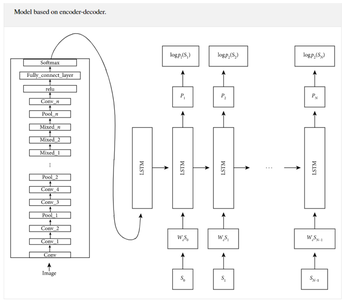
Implementation

- Show and Tell [1] model based on maximisation of probability of target sentence.

$$\theta^* = \arg max_\theta \sum_{I,S} \log p(S/I;\theta)$$

$$\log p(S|I) = \sum_{t=0}^{N} \log p([S_t/I, S_0, S_1, S_2.., S_{t-1}])$$

$$x_{-1} = \text{CNN}(I)$$
$$x_t = W_e S_t, \quad t \in \{0 \dots N-1\}$$
$$p_{t+1} = \text{LSTM}(x_t), \quad t \in \{0 \dots N-1\}$$



Results

Few examples of captions generated by the Show and Tell model.

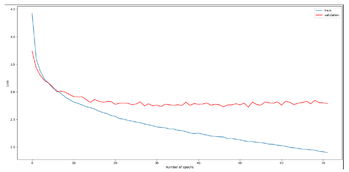
a group of people are standing in front of a crowd
Translated(src=en, dest=hi, text=तीन या एक समूह भीड़ के सामने खड़ा है, pronunciation=logon ka ek samooh bheed ke saamane khada hai

a black and white dog is standing on the grass
Translated(src=en, dest=hi, text=एक काला और सफेद कुत्ता घास पर खड़ा है, pronunciation=ek kaala aur saphed kutta ghaas par khada hai

a man on a bike is doing a trick on a dirt bike
Translated(src=en, dest=hi, text=बाइक पर एक आदमी गंदगी बाइक पर एक चाल चल रहा है

a little boy in a red shirt is sitting on a red slide
Translated(src=en, dest=hi, text=लाल शर्ट में एक छोटा लड़का लाल स्लाइड पर बैठा है,

CNN - Loss Resnet32, RNN – LSTM, Dataset – Flickr-8k



Performance Evaluation

Evaluated using BLEU [4],METEOR[6] , ROUGE[7] and CIDEr[5].

Model Implemented	Bleu1	Bleu2	Bleu3	Bleu4	ROUGE	METEOR	CIDEr
ResNet34 + LSTM	46.1	29.1	17.8	11.2	34.8	15.7	17.8

Future Goals

- Implementing visual attention based models.
- Use Indian languages based datasets to remove the error-margin of Google-trans.
- Work on Multilingual models.
- Getting more and more accurate with respect to human-annotated captions.

Refrences

[1] "Show and tell: A neural image caption generator." O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. , In CVPR, 2015
[2] "Neural machine translation by jointly learning to align and translate," K. Cho, D. Bahdanau, and Y. Bengio, Available at arXiv:1409.0473, 2014
[3] "K. Cho, H. Schwenk, B. van Merrienboer, F. Bougares, C. Gulcehre, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation," In EMNLP, 2014.,
[4] "BLEU: a method for automatic evaluation of machine translation," P. Kishore, S. Roukos, T. Ward, and W.-J. Zhu,, 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, USA, July 2002.
[5] "Cider: consensus-based image description evaluation," R. Vedantam, C. Lawrence Zitnick, and D. Parikh, IEEE Conference, pp. 4566–4575, Boston, MA, USA, June 2015.
[6] "METEOR: an automatic metric for MT evaluation with improved correlation with human judgments," S. Banerjee and L. Alon,, pp. 65–72, Ann Arbor, MI, USA, June 2005.
[7] "ROUGE: a package for automatic evaluation of summaries," C.-Y. Lin, Barcelona, Spain, July 2004.
[8] "Generating image descriptions using dependency relational patterns," A. Aker and R. Gaizauskas, In ACL, 2010.