For Research paper:

1. Structured followed: Efficient Urdu Caption Generation using Attention based LSTM.
2. Generative image captioning in Urdu using deep learning

Pages to Research.

1. Attention is all you need
2. R. Kiros, R. Salakhutdinov, and R. Zemel. Multimodal neural language models. In International conference on machine learning, pages 595–603, 2014.
3. A. Daud, W. Khan, and D. Che. Urdu language processing: a survey. Artificial Intelligence Review, 47(3):279–311, 2017.
4. R. Kiros, R. Salakhutdinov, and R. S. Zemel. Unifying visual-semantic embeddings with multimodal neural language models. arXiv preprint arXiv:1411.2539, 2
5. J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille. Deep captioning with multimodal recurrent neural networks (m-rnn). arXiv preprint arXiv:1412.6632, 2014
6. O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3156–3164, 201
7. K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning, pages 2048–2057, 2015.
8. J. A. Alzubi, R. Jain, P. Nagrath, S. Satapathy, S. Taneja, and P. Gupta, “Deep image captioning using an ensemble of cnn and lstm based deep neural networks,” Jan 2021
9. S. Herdade, A. Kappeler, K. Boakye, and J. Soares, “Image captioning: Transforming objects into words,” 2020

--------------------------------------------------------------------------------------------------------------------------------- .

***Urdu Image Captioning System:***

***Abstract:***

Image captioning is a process which involves generating human-like textual descriptions in the form of sentences or phrases to explain the content of an image. This technology is being widely applied to enhance accessibility for visually impaired individuals and improve user experience. Significant contributions have been made in image captioning, particularly in English. However, the field of image captioning in Urdu remains relatively unexplored. Urdu is the national language of Pakistan and also much spoken and understood in the subcontinent region of Pakistan-India, and yet a very little work has been done for Urdu language caption generation. In addition to this, the availability of dataset containing images of Pakistan which is a critical component for effective image captioning in Urdu, is nonexistent. Our research aims to fill this gap building an exclusive dataset featuring images of Pakistani roads, manually curated and sourced from the internet, with their captions annotated in Urdu. We will then apply various state-of-the-art Deep Learning models in order to perform image captioning on the proposed dataset, thereby addressing the unique linguistic and cultural context of Urdu language.

We aim to

(i) Present a new dataset featuring Roads of Northern Areas of Pakistan.

(ii) Present different architectures for image captioning in the Urdu language. These attention mechanisms are new to the Urdu language, as those have never been used for the Urdu image captioning.

---------------------------------------------------------------------------------------------------------.

***Literature review papers:***

1. ***Attention is all you need:***

- Introduces the Transformer model, a novel approach in the field of sequence modeling, particularly for tasks like machine translation.

- The Transformer model abandons the use of recurrent layers and relies entirely on an attention mechanism.

- Demonstrates state-of-the-art performance in English-to-German and English-to-French translation tasks, outperforming previous models.

- Utilizes self-attention, which allows the model to weigh the importance of different parts of the input sequence differently.

- The architecture consists of an encoder and a decoder, each composed of multiple identical layers.

- Each layer in the encoder contains two sub-layers: a multi-head self-attention mechanism and a position-wise fully connected feed-forward network.

- The decoder also has an additional third sub-layer, which performs multi-head attention over the output of the encoder stack.

- Employs positional encoding to give the model information about the relative or absolute position of the tokens in the sequence.

***2) Generative Urdu Image captioning:*** .

- It introduces a new dataset for Urdu image captioning, crucial for model training and evaluation.

-Explores various attention-based architectures for image captioning in Urdu, a novel approach for this language.

- A quantitative and qualitative analysis of different model architectures on Urdu image caption generation is performed.

- Data and code used in the study are made available for future research, fostering further advancements in this area.

***3) Efficient Urdu Caption Generation using Attention-based LSTM***.

- Aims to fill the gap in Urdu language caption generation by developing an attention-based deep learning model.

- Prepares an Urdu dataset by translating a subset of the "Flickr8k" dataset, containing 700 'man' images.

- Focuses on using sequence modeling techniques specialized for the Urdu language.

- Demonstrates the model's effectiveness with a BLEU score of 0.83 in Urdu.

- Enhances state-of-the-art models using improved CNN architectures and optimization techniques.

- Discusses the grammatical correctness of generated captions.

***Note:***

- Both papers address the under-researched area of Urdu language image captioning.

- The first paper introduces novel attention-based architectures and a new dataset, whereas the second paper focuses on improving existing models with better CNN architectures and an Urdu-translated subset of an existing dataset.

- The first paper provides a broader analysis with various BLEU scores, indicating a more extensive experimental setup.

- The second paper highlights the improvement in BLEU score achieved through model enhancement.

- Both contribute significantly to the field by providing resources and methodologies for Urdu image captioning, but with different approaches and focuses.