IMAGE CAPTION GENERATOR USING MACHINE LEARNING

Mr.K.PREMKUMAR, M.E, VLSI Design, B.E, Assistant Professor, Department of Computer science and Engineering,

Mr.M.LOKESH, B.E, Student of Computer Science Engineering

Mr.S.SYED IBRAHIM, B.E, Student of Computer science and Engineering

Mr.C.T. JERIN ROJER, B.E, Student of Computer science and Engineering

St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract

Image captioning, the task of automatically generating natural language descriptions for images, has garnered significant attention in the field of artificial intelligence. This project aims to develop an image caption generator using machine learning techniques, bridging the domains of computer vision and natural language processing .Feature extraction is performed using a pre-trained Convolutional Neural Network (CNN), which extracts high-level features from images. These features serve as input to a caption generation model, typically based on Recurrent Neural Networks (RNNs) or Transformer architectures. The model learns to generate captions by conditioning on both image features and previously generated words. Training the model involves optimizing parameters to minimize a suitable loss function, often based on maximizing the likelihood of generating ground truth captions given the input images. Evaluation metrics such as BLEU score, METEOR, and CIDEr are used to assess the quality of generated captions.

The project contributes to advancing the understanding of multimodal learning, where models learn to extract meaningful representations from both visual and textual modalities. Furthermore, the developed image caption generator has practical applications in various domains, including assistive technologies for the visually impaired, content recommendation systems, and image indexing and retrieval.

Key terms:

1. Image captioning 2. Machine learning

3. Convolutional Neural Network (CNN) 4. Recurrent Neural Network (RNN)

5. Transformer 6. Preprocessing

7. Feature extraction 8. Training

9. Inference 10. Evaluation metrics

11. Deployment

Introduction :

Image captioning, the task of automatically generating textual descriptions for images, is a fundamental problem at the intersection of computer vision and natural language processing. With the advancement of deep learning techniques, particularly convolutional neural networks (CNNs) for image processing and recurrent neural networks (RNNs) for sequence generation, significant progress has been made in developing robust image captioning systems. These systems have practical applications in various domains, including assistive technologies, content recommendation systems, and image indexing.

This project aims to explore the development of an image caption generator using machine learning techniques. By leveraging the power of CNNs to extract meaningful features from images and RNNs to generate coherent captions, we aim to build a model capable of accurately describing the contents of diverse images in natural language.

The project will involve dataset acquisition, preprocessing of images and captions, feature extraction using pre-trained CNN models, design and training of a caption generation model, evaluation of model performance using standard metrics, and deployment of the trained model for real-world use.

Through this project, we seek to contribute to the advancement of multimodal learning and demonstrate the practical utility of image captioning systems in various applications.

Literature survey

1. *"Show and Tell: A Neural Image Caption Generator" by Vinyals et al. (2015)*:

   - This seminal paper introduced the concept of using a CNN to encode images and an RNN to generate captions, laying the foundation for many subsequent image captioning models.

2. *"Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering" by Anderson et al. (2018)*:

   - This paper proposed an attention mechanism that allows the captioning model to focus on different regions of the image while generating captions, leading to improved performance.

3. *"DenseCap: Fully Convolutional Localization Networks for Dense Captioning" by Johnson et al. (2016)*:

   - DenseCap introduced a fully convolutional localization network for generating dense captions, providing descriptions for multiple regions within an image, which goes beyond traditional single-sentence captioning.

4. *"SCAN: Learning to Classify Images Without Labels" by Lee et al. (2019)*:

- This paper proposed a self-supervised approach for image captioning, where the model learns to generate captions by predicting image rotations. This eliminates the need for manually annotated captioning datasets.

5. *"Attention Is All You Need" by Vaswani et al. (2017)*:

- Although primarily focused on machine translation, this paper introduced the Transformer architecture, which has been successfully adapted for image captioning tasks, offering improvements in both efficiency and performance.

These papers provide a comprehensive overview of the advancements in image captioning research, from the initial approaches to the latest state-of-the-art techniques. They serve as foundational references for understanding the key concepts, architectures, and evaluation methodologies in the field.

System Design

1. *Data Acquisition and Preprocessing*:

- Acquire a dataset containing images paired with corresponding captions. Preprocess the images by resizing them to a uniform size and normalizing pixel values. Tokenize the captions into words or sub words and create a vocabulary mapping each word to a unique index. Pad or truncate captions to a fixed length.

2. *Feature Extraction*:

- Use a pre-trained Convolutional Neural Network (CNN), such as VGG16, Res-Net, or Inception, to extract features from the images. Remove the classification head of the CNN and use the output of the last convolutional layer as the image feature vector.

3. *Caption Generation Model*:

- Design a neural network architecture for generating captions. This typically involves an encoder-decoder framework.

- The encoder takes the image features as input and encodes them into a fixed-size vector representation.

- The decoder, often implemented as a Recurrent Neural Network (RNN) or Transformer, generates captions word by word based on the encoded image features.

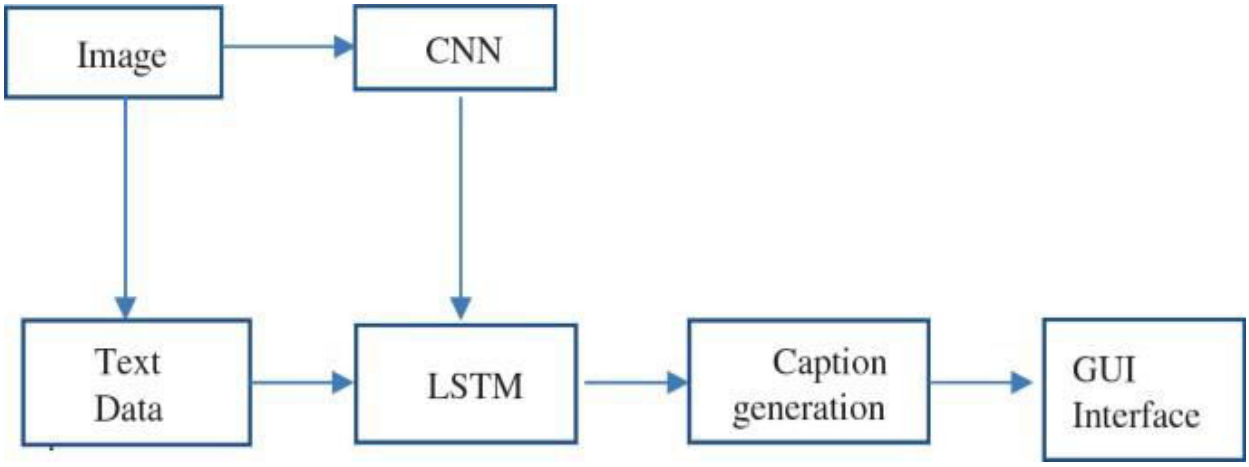
Image
CNN
Text Data
LSTM
Caption generation
GUI Interface

Snapshots


A group of people are playing a game of cards.

Conclusion

In conclusion, building an image caption generator using machine learning is a challenging but rewarding endeavor. By combining techniques from deep learning, computer vision, and natural language processing, developers can create models capable of generating descriptive captions for images. While the process involves training neural networks on large datasets and fine-tuning hyperparameters, the results can be impressive, opening up possibilities for applications in areas like accessibility, content indexing, and image understanding. As technology continues to advance, image caption generators may become even more accurate and versatile, offering new ways to interact with visual content in the digital world.

References

[1] Soheyla Amirian, Khaled Rasheed, Thiab R. Taha, Hamid R. Arabnia*§," Image Captioning with Generative Adversarial Network*", 2019 (CSCI).

[2] Miao Ma, Xi'an," *A Grey Relational Analysis based Evaluation Metric for Image Captioning and Video Captioning*". 2017 (IEEE).

[3] Niange Yu, *Xiaolin Hu, Binheng Song, Jian Yang, and Jianwei Zhang." Topic-Oriented Image Captioning Based on Order-Embedding*", (IEEE), 2019.

[4] Seung-Ho Han and Ho-Jin Choi" *Domain-Specific Image Caption Generator with Semantic Ontology*". 2020 (IEEE).

[5] Aghasi Poghosyan*," Long Short-Term Memory with ~~Read only Unit in~~ Neural Image Caption Generator*". 2017 (IEEE).

AUTHOR 1

Mr.PremKumar.K M.E, VLSI Design, Assistant Professor in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu.

AUTHOR 2

Mr. Syed Ibrahim. S B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I had attended many Workshops, Seminars in Python, Web development. I got placed in Reputed Company like Q Spider.

AUTHOR 3

Mr. Lokesh .M B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I had attended many Workshops, Seminars in Python, Web development.

AUTHOR 4

Mr. Jerin Rojer .C.T B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I had attended many Workshops, Seminars in Python.