1. Every Picture Tells a Story: Generating Sentences from Images: https://www.researchgate.net/publication/221303952_Every_Picture_Tells_a_Story_Generating_Sentences_from_Images

2. Automatic Generation of Descriptive Titles for Video Clips Using Deep Learning https://www.researchgate.net/publication/350750079_Automatic_Generation_of_Descriptive_Titles_for_Video_Clips_Using_Deep_Learning

3. The Use of Video Captioning For Fostering Physical Activity https://www.academia.edu/69366209/The_Use_of_Video_Captioning_for_Fostering_Physical_Activity (page: 3 encoder-decoder)

4. Key Clips and Key Frames Extraction of Videos Based on Deep Learning: https://iopscience.iop.org/article/10.1088/1742-6596/2025/1/012018/pdf
5. Self-Supervised Learning to Detect Key Frames in Videos: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7731244/
6. Show and Tell: A Neural Image Caption Generator: https://arxiv.org/pdf/1411.4555v2.pdf
7. Video Storytelling: Textual Summaries for Events https://arxiv.org/pdf/1807.09418.pdf

8. Deep Learning-Based Short Story Generation for an Image Using the Encoder-Decoder Structure https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9512087
9. https://www.academia.edu/74800815/Automatic_Generation_of_Descriptive_Titles_for_Video_Clips_Using_Deep_Learning- very important paper. Please have a look guys.

Eta emnei rakhsi for future reference**Dense-Captioning Events in Videos :https://openaccess.thecvf.com/content_ICCV_2017/papers/Krishna_Dense-Captioning_Events_in_ICCV_2017_paper.pdf

Algorithms/libraries:

1.Pixellib- Image segmentation and object detection
Requires: Tensorflow
2. Mask R-CNN- detects objects in an image and generates a high-quality segmentation mask for each instance

Related Projects:

1. Image to story
   https://github.com/ryankiros/neural-storyteller

How does it work?

1. Train a recurrent neural network (RNN) decoder on romance novels.
2. Each passage from a novel is mapped to a skip-thought vector.
3. Conditions RNN on skip-thought vector & generate the encoded passage.
4. Train a visual-semantic embedding between COCO images and captions.

   Captions and images are mapped into a common vector space.
5. After training, embed new images and retrieve captions.

Queries

1. Find a genre and make the model based on that?
2. To train deep learning models, high configured PC and GPU are needed sikeee
3. Key frame extraction vs no extraction

BEST ARTICLE:

https://medium.com/@samim/generating-stories-about-images-d163ba41e4ed