GramCam: Personalized Caption and Hashtag Generation for Social Media Posts

Abhey Kalia
IIIT Delhi
Delhi, India
abhey20420@iiitd.ac.in

Anirudh Ramasubramanian Iyer
IIIT Delhi
Delhi, India
anirudh20025@iiitd.ac.in

Bhavya Jain
IIIT Delhi
Delhi, India
bhavya20428@iiitd.ac.in

Karan Baboota
IIIT Delhi
Delhi, India
karan20210@iiitd.ac.in

Khwaish Rupani
IIIT Delhi
Delhi, India
khwaish20212@iiitd.ac.in

Nipun Gupta
IIIT Delhi
Delhi, India
nipun20089@iiitd.ac.in

ABSTRACT

The aim of this project is to develop a social media caption generator that can automatically generate engaging, sentimental and relevant captions and hashtags for images retrieved from social media platforms for the user's content in order to get a broad reach of the target audience. The pipeline we follow to achieve this is broken down into smaller tasks and the resulting successful models of each stage will be combined to achieve the final objective. In our work we first produce a descriptive caption from an image and then move on to making it more creative. The evaluation of our descriptive captions show comparable results with the existing state of the art. To evaluate the creative captions, we require human judgment but the results produced show significant improvement over the previous attempts to generate social media worthy captions.

MOTIVATION

The rise of social media in the past decade has been more than prominent. With its ever-growing user base comes the increasing requirement of those users to generate engaging and relevant captions. (Feng and Lapata, 2010) states that most search engines deployed on the web retrieve images without analyzing their content. They simply match user queries against collocated textual information like metadata, user-annotated tags (Eg. hashtags), captions, text surrounding the image. This limits the applicability of the search engines as images that do not coincide with the textual data cannot be retrieved. Thus, an appropriate caption-hashtag combination for users' images will help them and their businesses reach their target audience. By making the time-consuming and brain-racking process of caption writing generation effortless and personalizable, we will increase user engagement and mitigate the vulnerability of bad social media marketing influences.

PROBLEM FORMULATION

This research paper aims to address the problem of generating stylised, personalized, engaging and relevant captions for social media posts using deep-learning models. While descriptive captions generated by popular datasets like Flickr and COCO provide accurate information about an image, they may not be suitable for social media platforms, since we require personalized and stylized captions to capture the attention of users. The objective of this research is to develop a pipeline containing varied deep-learning models which take an image as an input and generate a stylised and personalised social media worthy caption along with hashtags and emojis.

The first step is to extract the salient features from the input image. These will be used to give direction to the descriptive caption as well as establish relevance between the caption generated from the image and the image itself.

Next, a deep learning model will be used to generate descriptive captions from the image which will then be passed into another deep learning model to transform it into a stylised caption while retaining the essence of the original sentence.

At this stage, we will inject the user's vocabulary to introduce a sense of personalisation into the caption.

Finally, frills like appropriate emojis and hashtags will be added to the caption to make it more relevant for social media and visually appealing.

We will employ pre-made datasets like Flickr, COCO and Conceptual Captions, as well as custom datasets of social-media-like captions for the above.

The proposed approach will enable businesses and individuals to improve their social media presence and resonate with their audience.

Github Repo: https://github.com/Nipun-Gupta26/IR-Project

Drive Folder:
https://drive.google.com/drive/folders/1XC9HkdsVrxp35ua70H9y8JDQv681Ysrc?usp=share_link

LITERATURE REVIEW

On reviewing multiple papers based on the concepts of caption generation we were able to identify the core areas that we should work on to make our model better than the previous research in the field.

Researches (Ramnath et al., 2014) focused on developing systems for generating captions for photos. This paper discussed two main approaches - The information retrieval approach which matches the image to a pre-existing database of captioned images and The generative computer vision approach which identifies the entities in the image and then generates a sentence. Their model relied on metadata to get information like camera model, time and location for context. However, it failed to mimic the human emotions in their captions.

(Kavi tha S et al., 2022) produced a caption generation model that employs CSPDenseNet and the FER model to extract salient image features and emotion features. It also introduced a self-attentive BiLSTM (bidirectional long short-term memory) network to process the textual knowledge and improve the quality of caption generation by focusing on the important text features as well as the contextual features.

(Park et al. 2017) also proposed a novel caption suggestion system which tackles the problem statement of personalized and accurate image captioning along with hashtag generation. They used CSMN (context sequence memory networks) to remember most frequently used words which introduced some level of personalisation. However, discrepancy persisted in some cases where the picture depicted a negative emotion while the model produced a positive caption. Moreover, the model had to be trained on each individual person instead of being able to incorporate user context in a scalable manner.

(Mathews et al., 2016) pointed out that in current caption generation methods, stylized non-factual descriptions are absent. They try to add emotion to the description generated as they are more likely to pique the interest of a reader. They introduced at least one emotion word whilst still keeping the essence of original description intact. They used CNN+RNN model. They used 2 parallel RNNs where one is a general background language model and the other specializes in generating descriptions with emotion. They also used word level regularizers to put more emphasis on sentiment words during training and combined the 2 RNN streams. In around 85% cases, their model was able to produce captions which are at least as descriptive as the factual captions. Though they are able to introduce sentiments into the generated captions, they are still factual and lack certain personalization. They also just consider 2 extreme emotions that are positive and negative.

The paper “Integration of textual cues for fine-grained image captioning using deep CNN and LSTM''. (Neeraj et al. 2019) suggests the fusion of text features present in an image with the semantic features to generate a more detailed caption. The authors suggest using the following model for achieving the results: use a fine-tuned visual geometry group (VGG) net for extracting semantic features, use a text saliency model to extract text features in the image where every word will be represented as a vector of 300 features, finally use long short-term memory (LSTM) for fusion. The authors have used Flickr8k and Flickr30k benchmark datasets for evaluating the model proposed. They have used BLEU and METEOR metrics to compare the proposed model with several other models, through which it is shown that this model outperforms the others. There are some problems associated with the proposed model where it is prone to false detection of texts due to false visual cues and because of exotic fonts.

Most of the models in this domain produce either non-sentimental captions which only take into account details like location and timestamp thereby making the captions impersonal and robotic or introduce some level of sentiment which may or may not appropriately correspond to the image. Some others try to provide captions using the same objects in different moods so as to be able to give a sentimental touch whereas others try to inculcate a personalised touch by storing user’s information like their relation to the people tagged in the post images or important dates like birthdays. However, none of the researchers tried to merge the tangents of stylisation and personalisation.

DATASETS

We used 3 benchmark datasets and 2 custom datasets at different stages of the pipeline.

The Flickr dataset is a sentence-based image description collection comprising of 8,000 images and five descriptive captions for each image which provide clear descriptions of the salient entities and events of the image.

The COCO (Common Objects in Context) dataset is a large-scale image recognition dataset for object detection, segmentation, and captioning tasks. It contains over 330,000 images, each annotated with 80 object categories and 5 captions describing the scene.

Google Conceptual captions is a dataset containing (image-URL, caption) pairs designed for the training and evaluation of machine learned image captioning systems. It consists of 2M+ images and their corresponding descriptive captions.

We randomly took 1000 elements from all three datasets comprising images and their captions to use them in our encoder. The encoder is responsible for generating descriptive captions from the images.

Next, we created our own Instagram dataset containing 6000+ images and their respective captions scraped from multiple public Instagram profiles. We used ChatGPT to generate a list of relevant Instagram public accounts. We specifically chose Instagram influencers having less than 200k followers to omit commercial posts and maintain variety. The prompt used is as follows:

“Give me names of 100 Instagram influencers' accounts with less than 200k followers, preferably no commercials in their feed, relevant captions that Instagram caption generator can use and variety in their feed not stuck to one topic”.

After getting the accounts, we used the Instaloader command line tool to scrape data from Instagram. We only scraped the posts from these accounts that were posted in 2023. We used this dataset for object extraction and user context.

Lastly, we also created our own creative captions dataset comprising 7000+ descriptive captions each having 3 creative insta-worthy captions along with hashtags and emojis. The descriptive captions for this were obtained from the conceptual captions dataset. For each (descriptive caption, creative caption 1, creative caption 2, creative caption 3) datapoint, we used chatGPT and manually gave it prompts to generate 3 insta-worthy captions. The prompt we used is as follows:

"Give me 3 instagram worthy captions of around 15-20 words for the given sentence. Add appropriate hashtags and emojis to the caption. Display as a dataframe with each caption in a seperate column. The first column will be the provided sentence. Don't change the sentence in the first column"

This dataset is used in training the model that generates Instagram relevant captions from descriptive captions.

4. METHODOLOGY

4.1 DATA PREPROCESSING

Instagram Scraped Dataset

For the data we had after scraping Instagram, we converted the captions into lower case, removed any hashtags or account tags present, removed punctuations. Dropped the accounts that had more than 30% of non-English words in their captions. Removed any non-English characters that were present if (<30%). Converted the captions into tokens using the NLTK library.

Creative Captions Dataset

For the creative captions dataset, we converted the (7305, 4) dataset into a (21915, 2) dataset. This was done by changing every (descriptive caption, creative caption 1, creative caption 2, creative caption 3) datapoint into (descriptive caption, creative caption) datapoint. This helped the model understand the different ways a descriptive caption can be turned into an stylised caption. After this, the NaN values were removed which led to the final dataset being (21915, 2) (Dataset did not have any NaN values)

This dataframe was then further divided into multiple other datasets. We extracted all the hashtags and emojis separately from the caption and stored it in their separate datasets. The caption without emojis and hashtags was stored in a (20098, 2) and (27855, 2) dataset respectively. In this way we could divide an Instagram-worthy caption into 3 main parts namely the caption text, hashtags and emojis and run models for each part separately.

4.2 FEATURE EXTRACTION

To generate a caption from an image, we need to extract the features from the image which can help describe the image. These features include image objects, scenes, and the contexts present in the image. These features act as queries which can help retrieve words which have been previously learnt by the model in a similar context. They can also be used to determine the relevance of the caption produced and rank them accordingly from most to least relevant.

For scene recognition, we deployed the pre-trained VGG16 model which has been trained on places365 and has 365 classes. We keep the top 2 predicted scenes in our query for the sake of accuracy.

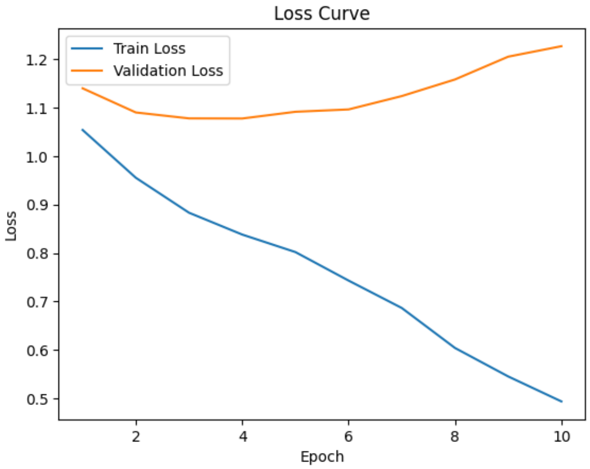
For the object detection task, we use the 'yolo5x' model which is able to predict objects from the 80 classes it has been trained on. This was done with the aim of getting the most relevant features out of the image and further using them to form meaningful sentences.

An example to demonstrate the above the methodology:
Query: Create a caption where the background is of either arena/performance or auditorium. The objects present in the image are: 15 persons, 1 backpack, 1 cell phone

The above query comprises the scenes and objects present in the image and uses it to generate a meaningful sentence. For this task, we fine-tune the T5 model on conceptual captions dataset where input is the encoded version of the above query and the target text is the caption from conceptual captions dataset.

Output: a large group of people in a field
Real Caption: a general view of confetti over the crowd

From the real caption, it can be seen that the scene and objects could be detected with a good precision and the model tried to use that information while generating a sentence but was not able to capture and take into account all the features in a meaningful way. This shows the limitation of this approach and the need to train it on more data so as to get desirable results.



Loss curve for 't-5 small' finetuning. The loss function used by the model is Cross-Entropy loss. From the loss curve, it can be inferred that the model was overfitting on the training data and more data is required to train the model and prevent this from happening.

4.3 USER CONTEXT

User context refers to the sentence structure and the common words that a user tends to use.

First, we extracted the context through TF-IDF scores of words used in the user corpus (i.e. set of captions by the user), similar to the approach used by (Park et al. 2017).
Our first approach was a simple word replacement algorithm where we extract top-k words from the user corpus wrt the TF-IDF scores. We then replace words in the generated sentence with similar words present in the top-k words through cosine similarity of words. The output for a synthetic user is as follows:

User top-5 words: {'cute': 0.45, 'sad': 0.2,'blue':0.15,'formula':0.15,'apple':0.05}
Input sentence: the dress looks beautiful and i love the lace.
Changed sentence: the dress looks cute and i love the lace.

It can be observed that’ beautiful’ in the original sentence has been replaced with ‘cute’ which is present in the top-5 words of the user. Though this approach does replace similar words, it didn't perform well with informal sentence structures where ‘hot’ is related to ‘beautiful’. In those cases, the cosine similarity of hot and beautiful wasn't enough to replace the word.

We then moved to implementing Logit Bias. Logit Bias makes amends to the word embeddings in a transformer based model so as to increase/decrease bias for certain words.
In the first approach to Logit Bias we increased the weights of the embeddings of certain words in the bert-base-uncased model with their TF-IDF scores:

User top-5 words: {'gorgeous': 0.85, 'sad': 0.15,'blue':0.15,'formula':0.1,'apple':0.05}
Changed bert-base output: the dress looks gorgeous... gorgeous. very gorgeous...... gorgeous................ gorgeous...................

It can be seen that the quality of the output even when the temperature of the model was adjusted is broken when the embeddings are messed around with. Thus, we hypothesized that this could be due to over-increase in the bias while making amends to the word embeddings. We then decreased the bias for similar words so when changes are made to the embedding matrix, words in the top-5 words will have a higher bias wrt similar words, but won’t be overpowering.
To find similar words we used "word2vec-google-news-300" which returned the top-50 similar words. While implementing this approach, we observed that ‘gorgeous’ was not being output while generic words like ‘amazing’ were being used which were not in the top-50 similar words to ‘gorgeous’. We further altered our method to increase the bias of the top-5 words by a small margin and decrease the relative bias of similar words. It can be seen that the entire class of words similar to ‘gorgeous’ are getting dropped when the bias is decreased while generic words are being used as fillers in the output. Just to be sure, we tested the same with "mrm8488/t5-base-finetuned-common_gen" model.

Changed bert-base output: the dress looks amazing.............
Prompt: her eyes were
Output for t5:
User top-5 words: {'cute': 0.45, 'sad': 0.2,'blue':0.15,'formula':0.15,'apple':0.05}
Input: her eyes were
Output: her eyes were a little blue and she was a little sad

At first the t5 model seemed to give good result with the above prompt but when tested with different prompts, the model was not making any difference than an unchanged model.

Input sentence: the beer was
Changed sentence: the beer was a little stale and the beer was a bit stale
Input sentence: she was the
Changed sentence: a woman was the only woman to ever be rescued

Finally our approach was to incorporate some information retrieval instead of completely changing embeddings for the model which seems scalable in theory. To do so, we collect the captions of a user and append them to a list. Then we apply preprocessing of the given set of captions (Stemming and Lemmatization) after which we calculate the top-k words through TF-IDF scores. After that, we simply append these words to the prompt. For example:

User captions: `["she looks hot in the new dress", "he was so absolutely cute today", "she looked absolutely slaying"]`
prompt: ‘mall person dress’
output: a person looks absolutely cute in a striped dress at a mall today
prompt: ‘dog, park, sad’
output: cute dog looking sad at the camera in the park slaying absolutly

It can be seen from the first prompt that the model is incorporating the top-5 words of the user but the sentence but when the prompt is changed to a different mood i.e. sad dog in a park, the caption generated doesn’t make any sense. This means that this approach is very mood specific where it captures the general mood of the top-5 words in the user corpus and forces it on every caption being generated.

Our final approach had a slightly better output than our previous approaches to change the embeddings linearly, which in turn is a non-linear problem that requires fine tuning through an optimizer or some machine learning problem which is not scalable.

4.4 PIPELINE DESCRIPTION

We propose a pipeline for end-to-end caption generation that takes an image as input and generates a stylized and relevant caption suitable for social media platforms such as Instagram.

Our pipeline consists of three major steps:

Descriptive Caption Generation: In the first phase of the pipeline, we use attention-based transformer models to generate a descriptive caption from the input image. We chose these models over LSTM and CNN models because of their ability to better capture relevant information. Specifically, we used the "nlpconnect/vit-gpt2-image-captioning" model, an encoder-decoder model trained on the Flick30k dataset, which encodes the image as input and generates the descriptive caption as output. This model was further fine-tuned on the COCO dataset to capture a variety of image features to retrieve the best possible words for the next phase of the pipeline.

Stylized Caption Generation: In the second phase of the pipeline, we generate a stylized caption from the descriptive caption using a sequence-to-sequence transformer model. We used the "mrm8488/t5-base-finetuned-common_gen" model, which is Google's T5 model fine-tuned on the CommonGen dataset, which generates sentences describing everyday scenarios using common sense reasoning. This makes this model suitable for our task as social media captions describe everyday photos of people in a fancy way. To fine-tune the model for our task, we added the prompt "write creatively: " to the descriptive caption, which was used as the input. The model was then trained to generate stylized captions suitable for social media by encoding the input and creative caption as the target text. Specifically, we fine-tuned the model on our creative caption dataset to learn how to rephrase the caption in an Instagram-like style.

Emoji and Hashtag Addition: In the final phase of the pipeline, we add emojis and hashtags to the stylized captions produced in the second phase, as they are essential components of social media captions. To accomplish this, we used the same sequence-to-sequence transformer model as in the second phase, but fine-tuned it separately on the emojis and hashtags datasets of our creative caption dataset. We trained the models using the input prompts "add emojis: " and "get tags: " for the emojis and hashtags, respectively. The target texts for these models were the emojis and hashtags, respectively.

Overall, our pipeline is able to generate an end-to-end Instagram caption that is both relevant and stylized for the given image, complete with emojis and hashtags. However, it is important to note that the pipeline may have limitations, such as overfitting to the training data and the need for large amounts of training data to ensure high-quality caption generation. Another limitation is that social media captions are subjective and a model might not produce captions to cater to everyone’s needs. We used three different instagram captions for each descriptive caption in the training process so that the model is able to take into account different ways expressing the same situation.

5. EVALUATION

We’re using 3 metrics to evaluate our model - BLEU score, METEOR score, ROGUE Score. Further, an essential metric of evaluation for our output is human feedback.

5.1 EVALUATION METRICS

BLEU stands for Bilingual Evaluation Understudy Score, and it is a metric used to evaluate a generated sentence with respect to a reference sentence, which in our case is the actual caption of the photo. It correlates highly with human evaluation. It works by counting the matching number of n-grams in the generated sentence to the n-grams in the reference sentence. The comparison is made regardless of the word order.

METEOR stands for Metric for Evaluation of Translation with Explicit Ordering. It captures more aspects of the generated sentence like word order, stemming and synonymy. It takes into account both precision and recall scores. It is calculated by taking the mean F-score computed with precision and recall, with recall weighted higher than precision.

While BLEU score is the most famous evaluation criteria for evaluating generated sentences, it has certain shortcomings which are fulfilled by the Meteor score. Meteor score presents a better correlation with human judgement.

ROGUE stands for Recall-Oriented Understudy for Gisting Evaluation. This metric compares the produced caption to the reference caption, in our case the original caption given by the dataset. It is computed as the ratio of the number of n-grams in reference that also appear also in produced caption over the number of n-grams in reference caption.

Human feedback is the best evaluation metric for our model since no pre-existing metrics cater to the definition of a social-media worthy caption. We applied NLP techniques like sentimental analysis, parts of speech evaluation, length of sentences to the descriptive and creative captions in an attempt to display the underlying differences between the two. However, none of the NLP techniques produced a metric for clear distinction between the descriptive and creative captions. Thus, human conscience is a better judge of the fact that a sentence with emojis, hashtags and flair is more appropriate as a social media caption as opposed to a descriptive sentence.

5.2 COMPARISON WITH BASELINE

Metric	**Baseline**	**Updated**
Bleu	0.5077187	0.570905
Meteor	0.3395625	0.3055678

*The above metrics are for the FlickR Dataset

5.3 PERFORMANCE ON EXISTING DATA

Metric	**FlickR**	**COCO**	**Conceptual**
Bleu	0.570905	0.74450008	1.160e-231
Meteor	0.3055678	0.27032845	0.2196645
Rouge	0.3507854	0.45333180	0.1203050

*Conceptual captions have the worst score for each metric as they're not exactly descriptive captions but somewhat close to a creative caption.

FlickR

Real caption - a person in gray stands alone on a structure outdoors in the dark

Generated caption - Snowy day vibes" ❤ #WinterStyle #MorningRoutine #CityVibes

COCO

Real caption - A man with a red helmet on a small moped on a dirt road.

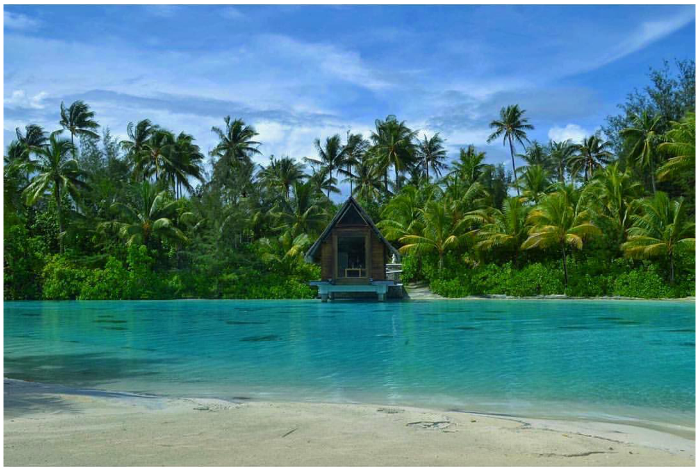
Generated caption - Riding through the dirt with my friends!"❤ #Motivation

Conceptual

Real Caption - the value of old olive trees

Generated caption - A tree with a lot of leaves" 🎨 #Haped Feelings

5.4 PERFORMANCE ON NEW DATA



Insta caption: 📍Bora Bora | Follow TravelTuesday to discover more places ✈️ #traveltuesday

Descriptive caption: a boat is docked in the water near a beach

Final generated caption: Seaside tranquility 🌊 #SummerVibes #SunsetScenes #Vacatio

5.5 SoTA

FlickR dataset

Literature	BRNN	Cornia et. Al	Unified VLP	Our model
Bleu Score	0.157	0.213	0.301	0.57

*Source: https://paperswithcode.com/sota/image-captioning-on-flickr30k-captions-test

* the reported baseline are on the BLEU-4 and our evaluation is on average BLEU score so direct comparison can not be made on these values

COCO dataset

Literature	LEMON	OFA	mPLUG	Our model
Bleu Score	0.426	0.45	0.465	0.744

*Source: https://paperswithcode.com/sota/image-captioning-on-coco-captions

6. NOVELTY

The novelty of our research lies in the creation of a custom dataset for generating creative social media-like captions. In contrast to existing datasets that focus on factual descriptions, our custom dataset included 3 creative captions for the factual description without changing its essence. We used this custom dataset to train a model that could convert descriptive captions into social media-like captions which are engaging. Our approach offers a novel and effective solution for generating engaging captions that are suitable for social media platforms. The ability to generate captions that are both descriptive and engaging has numerous applications in areas such as marketing, advertising, and social media management.

Furthermore, our attempt to collect diverse captions from a pool of participants with varying backgrounds and interests in a scalable manner yielded a model that could incorporate user-context to some extent. This could not be merged with the main pipeline due to the lack of meaning when there is a change in mood of the caption. However, this opens up possibility for future work in this domain.

7. FUTURE WORK

With an aim to make the caption personalised to the user's preferences, the user context tangent of the research can be refined and integrated with the main pipeline. Future work can also be focused on capturing the sentence structure of a certain user, i.e. following the same grammatical pattern a user makes in their set of captions.

The method proposed using the scene and object detection shows some promising results but the captions generated were not able to capture the image properly. Further improvements in this approach may help make the captions better as more information can be retrieved using these features making the captions more relevant.

Further, the current emoji generator model is a classifier. Due to data limitations, the classification is limited to a one emoji for four to five keywords. A more extensive dataset will lead to better classification as well as multiple emoji options per keyword.

An Information Retrieval approach would be to offer a choice of captions to the user to choose from and determine engagement metrics. These can be utilised to generate rankings of future captions produced on the basis of the user's current feedback. Thus, the model can be customised to the user's caption preferences.

8. INDIVIDUAL TASKS

Anirudh R. Iyer	User context, Report, PPT
Abhey Kalia	Feature extraction, Report, PPT
Bhavya Jain	Creative Captions dataset, Instagram Handles dataset, report, PPT
Karan Baboota	Descriptive Caption generation, Hashtag generation, Report, PPT
Khwaish Rupani	Creative Captions dataset, Creative Captions generation, Report, PPT
Nipun Gupta	Creative Captions generation, Emoji generation, Report, PPT

REFERENCES

[1] Ramnath, K. et al. (2014) "Autocaption: Automatic Caption Generation for personal photos," IEEE Winter Conference on Applications of Computer Vision [Preprint]. Available at: https://doi.org/10.1109/wacv.2014.6835988.

[2] Park, C.C., Kim, B. and Kim, G. (2017) "Attend to you: Personalized image captioning with context sequence memory networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Preprint]. Available at: https://doi.org/10.1109/cvpr.2017.681.

[3] Kavi tha S, Pon Karthika K, Jayakumar Kaliappan, Selvaraj, S., R. Nagalakshmi and Molla, B. (2022). Caption Generation Based on Emotions Using CSPDenseNet and BiLSTM with Self-Attention. Applied Computational Intelligence and Soft Computing, 2022, pp.1–13. doi:https://doi.org/10.1155/2022/2756396.

[4] Gupta, N. and Jalal, A.S. (2019) "Integration of textual cues for fine-grained image captioning using deep CNN and LSTM," Neural Computing and Applications, 32(24), pp. 17899–17908. Available at: https://doi.org/10.1007/s00521-019-04515-z.

[5] Mathews, A., Xie, L. and He, X. (2016) "Senticap: Generating image descriptions with sentiments," Proceedings of the AAAI Conference on Artificial Intelligence, 30(1). Available at: https://doi.org/10.1609/aaai.v30i1.10475.

[6] Feng, Y. and Lapata, M. (2010). How Many Words Is a Picture Worth? Automatic Caption Generation for News Images. Meeting of the Association for Computational Linguistics, pp.1239–1249.