Database, Data Mining, Machine Learning, Natural Language Processing

[deep learning](https://scholar.google.com/citations?view_op=search_authors&hl=en&mauthors=label:deep_learning) [artificial intelligence](https://scholar.google.com/citations?view_op=search_authors&hl=en&mauthors=label:artificial_intelligence)

* Youtube
* Google
* Calls sr shan
* Gpt

Upwork

Text-to-video

Visual ques answering

Fake News Detection and Misinformation Analysis

**1. Text-to-Video Generation**

**Problem Statement**: Enable easy creation of high-quality videos from text for educators and content creators.

**2. Visual Question Answering (VQA)**

**Problem Statement**: Assist visually impaired and students by answering questions about image content accurately.

**3. Fake News Detection and Misinformation Analysis**

**Problem Statement**: Detect fake news and analyze misinformation spread on social media to ensure credible information.

**1. Text-to-Video Generation**

**Problem Statement**: Automate video creation from text for non-technical users in education and marketing.

**2. Visual Question Answering (VQA)**

**Problem Statement**: Provide accurate image-based question answering for enhanced accessibility and educational tools.

**3. Fake News Detection and Misinformation Analysis**

**Problem Statement**: Identify fake news and track misinformation on social media to improve information integrity.

visualbert

<https://huggingface.co/docs/transformers/en/model_doc/visual_bert>

<https://github.com/uclanlp/visualbert>

<https://github.com/kHarshit/visual-question-answering>

<https://medium.com/data-science-at-microsoft/visual-question-answering-with-multimodal-transformers-d4f57950c867>

<https://youtu.be/DkzbCJtFvqM?si=SyCUs-8kBKLxfGzT>

<https://youtu.be/V1-Hm2rNkik?si=k9O9Lep6ZsGnj8Md>

<https://medium.com/@niralidedaniya/visual-question-answering-a-deep-learning-classification-case-study-3d8df1eb48e3>

<https://jalammar.github.io/illustrated-bert/>

<https://www.javatpoint.com/multimodal-transformer-models#:~:text=Multimodal%20transformer%20models%20extend%20the,image%20captioning%2C%20and%20speech%20recognition>.

<https://getliner.com/search/s/19781662/t/64051305>

Problems

Gpu

Site

personalized

S. I. Khandaker, T. Talukdar, P. Sarker, M. H. K. Mehedi, E. R. Rhythm and A. A. Rasel, "Vision Meets Language: Multimodal Transformers Elevating Predictive Power in Visual Question Answering," 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICCIT60459.2023.10441514. keywords: {Visualization;Image coding;Computational modeling;Medical services;Predictive models;Transformers;Question answering (information retrieval);Visual Question Answering (VQA);Benchmark Datasets;Multimodal Transformers;Interpretability},

S. Ravi, A. Chinchure, L. Sigal, R. Liao and V. Shwartz, "VLC-BERT: Visual Question Answering with Contextualized Commonsense Knowledge," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 1155-1165, doi: 10.1109/WACV56688.2023.00121. keywords: {Comets;Visualization;Analytical models;Knowledge based systems;Linguistics;Transformers;Question answering (information retrieval);Algorithms: Vision + language and/or other modalities;Image recognition and understanding (object detection;categorization;segmentation;scene modeling;visual reasoning)},

M. Dias, H. Aloj, N. Ninan and D. Koshti, "BERT based Multiple Parallel Co-attention Model for Visual Question Answering," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1531-1537, doi: 10.1109/ICICCS53718.2022.9788253. keywords: {Training;Visualization;Computational modeling;Bit error rate;Natural languages;Predictive models;Feature extraction;Visual Question Answering;Image Question answering;Hierarchical VQA;BERT},

D. Koshti, A. Gupta and M. Kalla, "Knowledge Blended Open Domain Visual Question Answering using Transformer," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 823-828, doi: 10.1109/ICAIS56108.2023.10073911. keywords: {Visualization;Knowledge based systems;Natural languages;Transformers;Question answering (information retrieval);Commonsense reasoning;Image question answering;Bidirectional Encoder Representation from Transformers;Visual question answering;ConceptNet;Knowledge visual question answering},

Z. Shao et al., "Visual Explanation for Open-Domain Question Answering With BERT," in IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 7, pp. 3779-3797, July 2024, doi: 10.1109/TVCG.2023.3243676.

keywords: {Analytical models;Transformers;Task analysis;Data models;Visual analytics;Bit error rate;Semantics;Explainable machine learning;open-domain question answering;visual analytics},

Z. Yang, N. Garcia, C. Chu, M. Otani, Y. Nakashima and H. Takemura, "BERT Representations for Video Question Answering," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 1545-1554, doi: 10.1109/WACV45572.2020.9093596. keywords: {Visualization;Bit error rate;Feature extraction;Knowledge discovery;Task analysis;Semantics;Standards},

M. Dias, H. Aloj, N. Ninan and D. Koshti, "BERT based Multiple Parallel Co-attention Model for Visual Question Answering," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1531-1537, doi: 10.1109/ICICCS53718.2022.9788253. keywords: {Training;Visualization;Computational modeling;Bit error rate;Natural languages;Predictive models;Feature extraction;Visual Question Answering;Image Question answering;Hierarchical VQA;BERT},

N. Akhila, S. J. P and S. K. P, "Comparative Study of Bert Models and Roberta in Transformer based Question Answering," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-5, doi: 10.1109/CONIT59222.2023.10205622. keywords: {Training;Measurement;Analytical models;Transfer learning;Bit error rate;Transformers;Data models;Natural Language Processing;Bert Models;SQuAD;question-answering;transformer},

D. Amin, S. Govilkar and S. Kulkarni, "Visual Question Answering System for Indian Regional Languages," 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 22-27, doi: 10.1109/ICAST55766.2022.10039528. keywords: {Deep learning;Visualization;Natural languages;Benchmark testing;Question answering (information retrieval);visual question answering;question answering},

F. Wang and G. An, "Visual Question Answering based on multimodal triplet knowledge accumuation," 2022 16th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2022, pp. 81-84, doi: 10.1109/ICSP56322.2022.9965282. keywords: {Visualization;Knowledge based systems;Semantics;Euclidean distance;Signal processing;Predictive models;Question answering (information retrieval);Visual Question Answer;Multimodal knowledge represent;Explicit triplet},

Project start

Research papers nd choose  
Env setup

Library installation

Comparison of datasets

Jason convert

Data Loading

Json to csv

import torch

from torch import nn

from torch.utils.data import DataLoader, Dataset

from transformers import BertTokenizer, ViTFeatureExtractor

class VQAModel(nn.Module):

def \_\_init\_\_(self):

super(VQAModel, self).\_\_init\_\_()

self.text\_model = BertModel.from\_pretrained('bert-base-uncased')

self.image\_model = ViTModel.from\_pretrained('google/vit-base-patch16-224-in21k')

self.classifier = nn.Linear(self.text\_model.config.hidden\_size + self.image\_model.config.hidden\_size, 1000) # Example

def forward(self, input\_ids, attention\_mask, pixel\_values):

text\_features = self.text\_model(input\_ids, attention\_mask=attention\_mask).pooler\_output

image\_features = self.image\_model(pixel\_values).pooler\_output

combined\_features = torch.cat((text\_features, image\_features), dim=1)

output = self.classifier(combined\_features)

return output

class VQADataset(Dataset):

def \_\_init\_\_(self, data, tokenizer, feature\_extractor):

self.data = data

self.tokenizer = tokenizer

self.feature\_extractor = feature\_extractor

def \_\_len\_\_(self):

return len(self.data)

def \_\_getitem\_\_(self, idx):

item = self.data.iloc[idx]

question = item['question']

image\_path = item['image\_path']

answer = item['answer']

image = Image.open(image\_path).convert("RGB")

inputs = self.tokenizer(question, return\_tensors="pt")

pixel\_values = self.feature\_extractor(images=image, return\_tensors="pt").pixel\_values

return {

'input\_ids': inputs['input\_ids'].squeeze(),

'attention\_mask': inputs['attention\_mask'].squeeze(),

'pixel\_values': pixel\_values.squeeze(),

'labels': torch.tensor(answer)

}

def main():

tokenizer = BertTokenizer.from\_pretrained('bert-base-uncased')

feature\_extractor = ViTFeatureExtractor.from\_pretrained('google/vit-base-patch16-224-in21k')

train\_df = pd.read\_csv('train\_data.csv')

train\_dataset = VQADataset(train\_df, tokenizer, feature\_extractor)

train\_loader = DataLoader(train\_dataset, batch\_size=8, shuffle=True)

model = VQAModel()

optimizer = torch.optim.Adam(model.parameters(), lr=1e-4)

criterion = nn.CrossEntropyLoss()

for epoch in range(10):

model.train()

for batch in train\_loader:

optimizer.zero\_grad()

outputs = model(batch['input\_ids'], batch['attention\_mask'], batch['pixel\_values'])

loss = criterion(outputs, batch['labels'])

loss.backward()

optimizer.step()

print(f'Epoch {epoch}, Loss: {loss.item()}')

torch.save(model.state\_dict(), 'vqa\_model.pth')

if \_\_name\_\_ == "\_\_main\_\_":

main()