



أكاديمية سدايا
SDAIA Academy



IMAGE CAPTIONS

IMAGE CAPTIONS

Deep Learning Project

- | | |
|---------------------|-------------------|
| • Batoul Alosaimi | • Amal Altamran |
| • Alanoud Alhussain | • Amirah Alotaibi |
| • Norah AlQahtani | • Shroaq Almutiri |
| • Asma Alsulami | |

An abstract background on the left side of the slide featuring a complex network graph with numerous nodes and connecting lines, creating a web-like structure.

Algorithms and Tools

- Keras
- TensorFlow
- PIL
- collections, random, re
- Convolutional Neural Networks
- Pandas, Numpy
- matplotlib, seaborn
- pydotplus
- googletrans
- gTTS



Introduction

Methodology

Conclusion

Data Set

Result

Future
Work

Image Captioning (IC)

Objective:

- Help disabilities
- Facilitate education

Prediction Caption: a group of young people are standing in a blurry near a wall



Out[83]:



Our Class 😊

Dataset:



Flick8k_Dataset :- contains more than 8000 images

- Contains the image id along with the 5 captions in the English language.**
- Contains the image id along with the 3 captions in the Arabic language.**



Methodology

Load
Data

Pre-
Processing

EDA

Split
Data

Transfer
Learning

Model

Evaluation

Exploratory Data Analysis



فتاة صغيرة مغطاة بالطلاء تجلس أمام قوس قزح
فتاة صغيرة تجلس أمام قوس قزح ملون كبير
فتاة أمام لوحة قوس قزح

Figure 6: image from Arabic data set

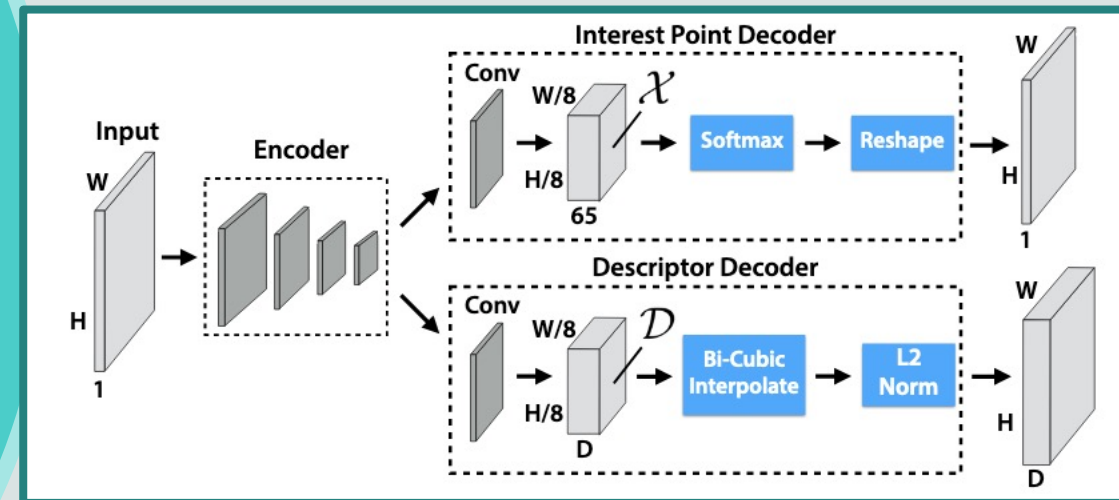


A little girl covered in paint sits in front of a painted rainbow with her hands in a bowl .
A little girl is sitting in front of a large painted rainbow .
A small girl in the grass plays with fingerprints in front of a white canvas with a rainbow on it .
There is a girl with pigtails sitting in front of a rainbow painting .
Young girl with pigtails painting outside in the grass .

Figure 7: image from English data set

Model used: Transfer Learning

Inception V3: Is a Convolutional Neural Network for Assisting in Image Analysis and Object Detection.



Using four Experiment:

- Arabic : we tried two different Epochs 8, 15
- English : we tried two different Epochs 8, 15

Model Evaluation

BLEU – Score:

(BiLingual Evaluation Understudy) evaluating machine-translated text

Audio Libraries

```
!pip install googletrans  
!pip install gTTS
```

gTTS Google Text to Speech, it is a Python library to interface with Google Translate text to speech API.

Note : This slide is attached to the audio

Result



Prediction Caption: طفل يرفع ذراعيه بالقرب من المعطر في الشارع

لفط	عفري	هوعارذ	برقالب
نم	رطملا	يف	عراشلا
<end>			

figure8: predicted image caption by the Arabic model

Prediction Caption: لاعب كرة قدم يستعد لتحمي الكرة

بغال	فرك	موق
دعنتسي	بيوصتل	فركلا
<end>		

Figure 5: Predicted Arabic caption for Saudi football Player Mohammed Noor picture from the internet



Figure: predict image caption by the English model



Real Caption: Group of people at party.
Prediction Caption: people dancing in the street

people	dancing	in
the	street	<end>

One[122]

Figure: predict image caption by the English model



Conclusion



- We created a model that predict image description (caption) so people who cannot see will be able to know what inside images
- In addition, we improved a program that convert caption text into voice.
- Our project support both Arabic and English languages.

Future Work



Work on a larger dataset that contain local images

Improve Arabic voice model

Create website



Thank You

✓ Batoul Alosaimi

✓ Alanoud Alhussain

✓ Norah AlQahtani

✓ Asma Alsulami

✓ Amal Altamran

✓ Amirah Alotaibi

✓ Shroaq Almutiri



IMAGE CAPTIONS