Image Caption

Generator

Redes Neuronales para el Análisis de Series Temporales

Diplomatura de Especialización en Inteligencia Artificial

Septiembre 2020

Integrantes...

Jorge Rodríguez Castillo

Keven Fernandez Carrillo

David Fosca Gamarra

Image Captioning

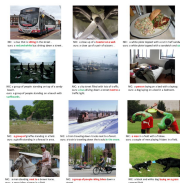**Caption:** Two dogs are playing on the snow

Caption: A basketball player with white uniform is dribbling the ball
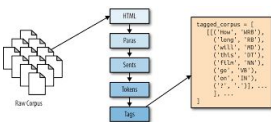
Solución a problemas reales...

Metodología

Paso 1

Selección de Dataset

Paso 2

Pre-procesamiento de Captions.

Paso 3

Generación de Image Feature Vectors.

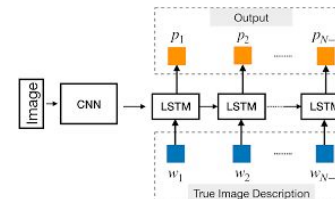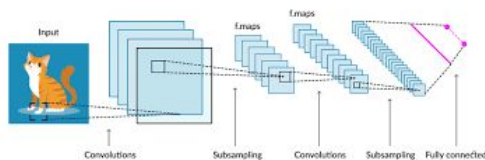**Paso 4**

Generación de secuencias para entrenamiento

Paso 5

Definición de Arquitectura.

Paso 6

Pruebas y resultados

Metodología

1. Selección de DataSet

Flickr8K
1.14 GB

Inglés
Idioma

8091
imágenes

40,455
captions

5 captions / imagen
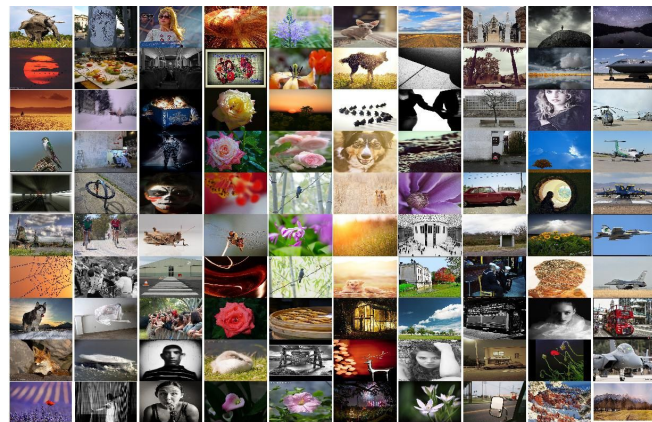
Duplicados/Sin texto

34 palabras
Longitud más grande
de caption

2232 palabras

Tamaño relevante de
vocabulario (> 1 palabra)

Metodología

1. **Selección de DataSet**

A little girl in a pink dress going into a wooden cabin .

A little girl climbing the stairs to her playhouse .

A little girl climbing into a wooden playhouse .

A girl going into a wooden building .

A child in a pink dress is climbing up a set of stairs in an entry way .
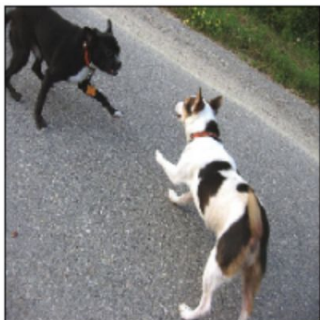
Two dogs on pavement moving toward each other .

Two dogs of different breeds looking at each other on the road .

A black dog and a white dog with brown spots are staring at each other in the street .

A black dog and a tri-colored dog playing with each other on the road .

A black dog and a spotted dog are fighting

5 captions / imagen

Metodología

2. Pre-procesamiento de Captions 🧹

- Generamos tokens.
- Eliminamos tokens de menos de 2 caracteres.
- Convertimos tokens a letras minúsculas.
- Eliminamos tokens numéricos.
- Concatenamos los tokens para formar nuevamente textos.
- Agregamos inicio y fin de secuencia a los textos.

```
['startseq dog is shaking off water endseq',
 'startseq dog shakes his head endseq',
 'startseq dog shakes off water endseq',
 'startseq dog shakes off water and gets contorted endseq',
 'startseq the dog is shaking water off of his body endseq']
```
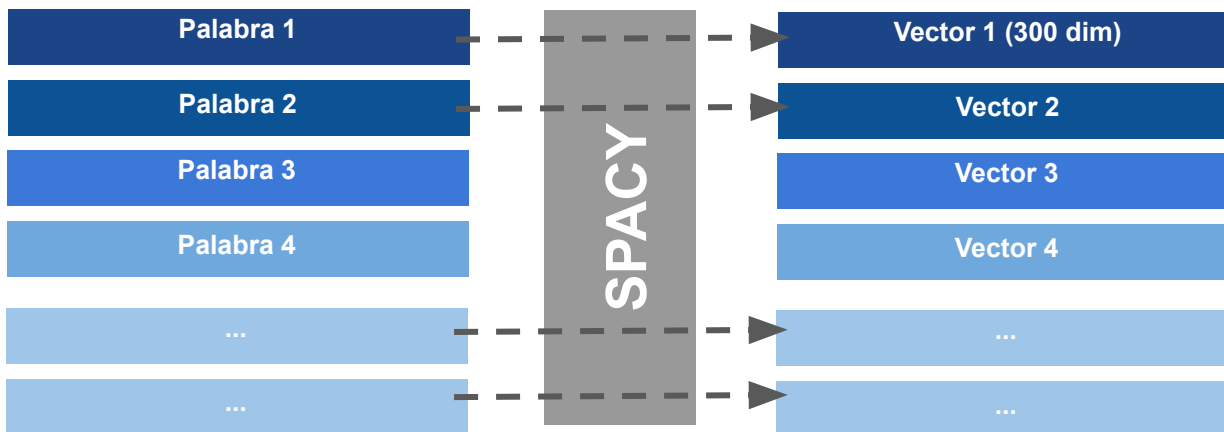
Metodología

2. Pre-procesamiento de Captions

- Generamos **matriz de vectores** en base a la palabras del vocabulario relevante y el Word Embedding empleado.

```
nlp = spacy.load("en_core_web_md")
```

Metodología

3. Generación de Feature Vectors

InceptionV3
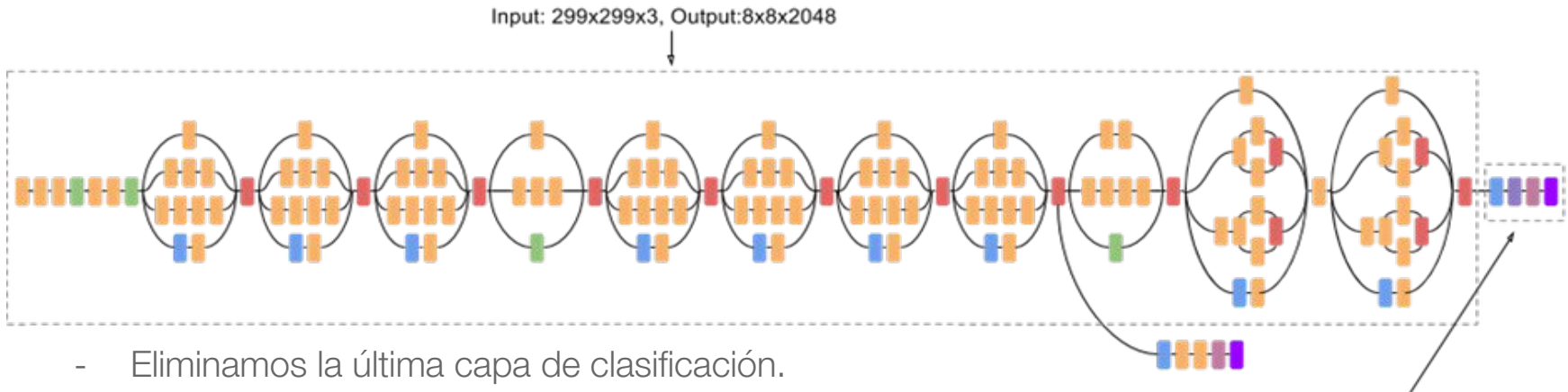
Input: 299x299x3, Output:8x8x2048

Final part:8x8x2048 -> 1001

- Eliminamos la última capa de clasificación.
- Resultando finalmente con vectores de 2048.
- Almacenamos vectores en ".pkl".

```
(array([0.25931156, 0.49846658, 0.21638362, ..., 0.12265647, 0.02689273,
       0.31099927], dtype=float32), (2048,))
```

Metodología

3. Generación de Feature Vectors

ResNet50

- Eliminamos la última capa de clasificación.
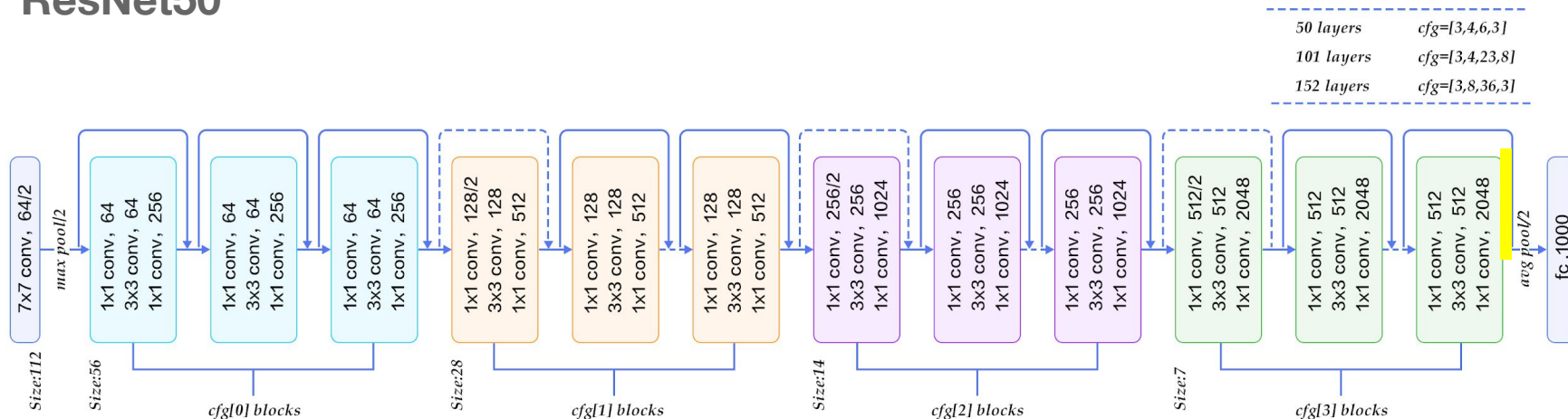- Resultando finalmente con vectores de 2048.
- Almacenamos vectores en ".pkl".

Metodología

4. Generación de secuencias para entrenamiento

Caption: Two dogs are playing on the snow

<startseq> Two dogs are playing on the snow **<endseq>**

Dividimos la secuencia en tramos

img_123456.jpg

Xcaption (i)	Y (i)
<startseq>	two
<startseq> Two	dogs
<startseq> Two dogs	are
<startseq> Two dogs are	playing
<startseq> Two dogs are playing	on
<startseq> Two dogs are playing on	the
<startseq> Two dogs are playing on the	snow
<startseq> Two dogs are playing on the snow	<endseq>

Metodología

4. Generación de secuencias para entrenamiento

Caption: Two dogs are playing on the snow

<startseq> Two dogs are playing on the snow **<endseq>**

word to index

img_123456.jpg

Xcaption (i)	Y (i)
9	7
9, 7	2
9, 7, 2	1
9, 7, 2, 1	4
9, 7, 2, 1, 4	3
9, 7, 2, 1, 4, 3	6
9, 7, 2, 1, 4, 3, 6,	5
9, 7, 2, 1, 4, 3, 6, 5	8

Metodología

4. Generación de secuencias para entrenamiento

Caption: Two dogs are playing on the snow

<startseq> Two dogs are playing on the snow **<endseq>**

zero's padding

Xcaption (i)	Y (i)
9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	7
9, 7, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	2
9, 7, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	1
9, 7, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	4
9, 7, 2, 1, 4, 0, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	3
9, 7, 2, 1, 4, 3, 0, 0, 0, 0, 0, 0, 0, ... , 0 , 0	6
9, 7, 2, 1, 4, 3, 6, 0, 0, 0, 0, 0, 0, ... , 0 , 0	5
9, 7, 2, 1, 4, 3, 6, 5, 0, 0, 0, 0, 0, ... , 0 , 0	8

img_123456.jpg

Metodología

5. Definición de Arquitectura
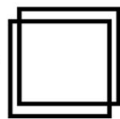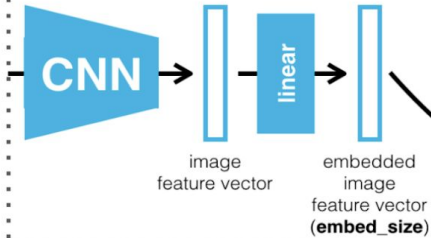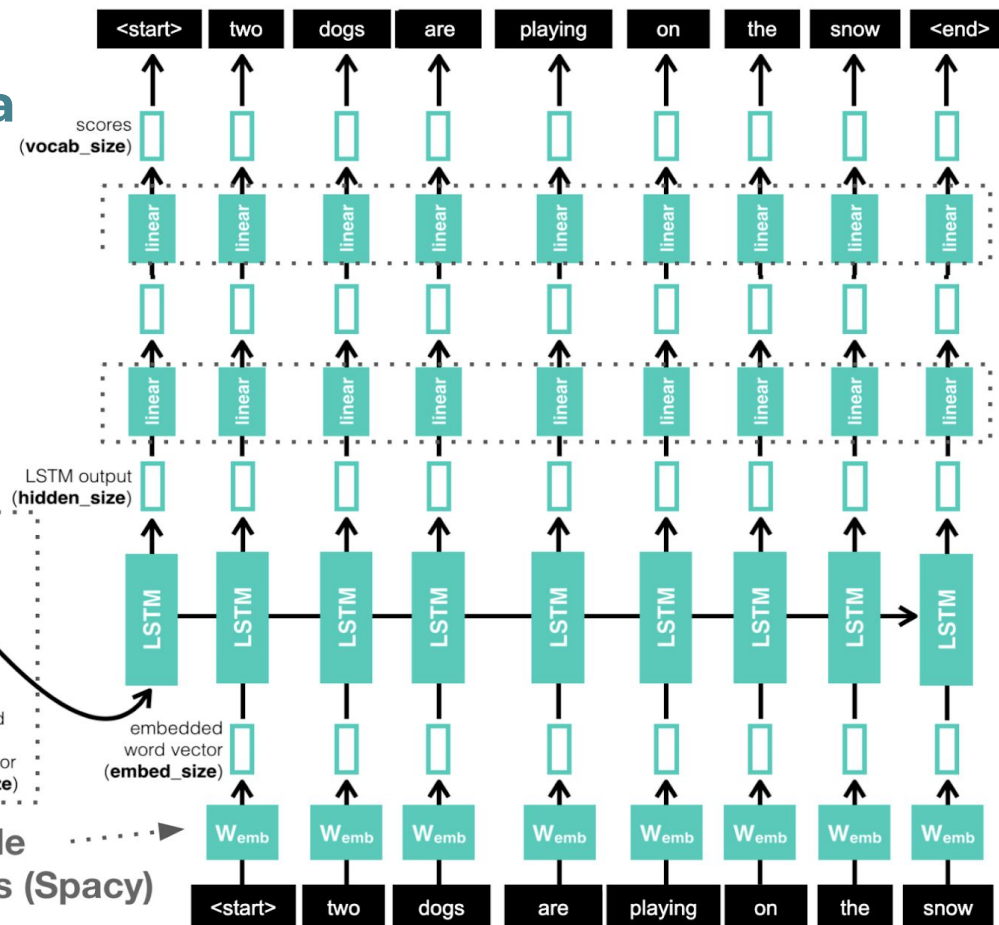
Metodología

5. Entrenamiento

50 épocas

26 min - GPU

Adam

Optimizador

32 imágenes

Batch Size

categorical_crossentropy

Función de Loss

0.0012 lr

Learning rate

Loss & Accuracy:

```
loss: 2.6940 - accuracy: 0.3657
```

Metodología

6. Pruebas y resultados

Caption 1: bald man dribbles basketball while wearing miami jersey
Caption 2: man playing basketball jersey number
Caption 3: number five plays basketball for miami
Caption 4: the basketball player is wearing orange sneakers
Caption 5: the number five miami player dribbles down the court

Prediction: basketball player in white uniform is dribbling the ball

Caption 1: couple people running through large sidewalk fountain
Caption 2: young boy is running through water fountain
Caption 3: children are under large fountain in the city
Caption 4: child runs through city fountain
Caption 5: some children run through fountain

Prediction: two children are playing in the water

Metodología

6. **Pruebas y resultados**

Caption 1: bicyclist rides on ramps
Caption 2: bike rider does trick in the air
Caption 3: man performs bike trick on ramp
Caption 4: extreme bmx rider defying gravity
Caption 5: person doing tricks on bicycle in skatepark

Prediction: man does trick on his bike in the air

Caption 1: woman in is walking holding silver cellphone
Caption 2: woman in black shirt walking down the street
Caption 3: woman wearing black shirt walks among crowd in city
Caption 4: brunette woman in black walking through blurred crowd
Caption 5: lady walking and squinting from the sun in large group of people

Prediction: woman in black shirt and sunglasses is standing in front of crowd

Metodología

6. Pruebas y resultados

Caption 1: two brown dogs are playing with each other in the snow
Caption 2: two brown dogs chase each other in the snow
Caption 3: two brown dogs play in the snow
Caption 4: two dogs playing in the snow
Caption 5: two dogs playing or fighting in the snow

Prediction: two dogs are running through the snow

Caption 1: guy is watching another guy jump in the air while snowboarding
Caption 2: person in orange pants is doing tricks on snowboard while
Caption 3: person sits on snowbank watching snowboarder take high leap
Caption 4: person watches as snowboarder does fancy trick in the air
Caption 5: snowboarder performing trick high in the air in snowy area

Prediction: snowboarder is jumping over snow hill

Metodología

6. Pruebas y resultados

Caption 1: dog chases toy in the grass with its owner in the background
Caption 2: dog chases white animal on green lawn
Caption 3: dog playing outside in the grass
Caption 4: golden dog is running across the grass chasing white toy on the ground
Caption 5: the brown dog is running outside on the grass

Prediction: dog runs through the grass

Caption 1: bunch of people playing soccer
Caption 2: action photo of soccer players in red and black uniforms
Caption 3: soccer player jumping in air during game
Caption 4: six soccer players on field with player in red uniform in the air and ball airborne
Caption 5: the red team knocked the ball toward the goal and the black team tried to block it

Prediction: two men in red and white uniforms play soccer

Metodología

6. Pruebas y resultados

Caption 1: little boy walking on railroad tracks
Caption 2: little boy walks along railroad tracks in the woods
Caption 3: little boy walking with stick on tracks
Caption 4: the little boy walks along the railroad tracks with his stick
Caption 5: there is little boy in red jacket walking down the railroad tracks
Prediction: man in red shirt is standing on the sidewalk

Caption 1: one girl is wearing white with red sleeves the other has words on hers
Caption 2: two asian girls laughing together
Caption 3: two girls with dark hair and white shirts
Caption 4: two girls with long hair are smiling and one is hold something white and green
Caption 5: two women in white shirts talking
Prediction: woman in black shirt and sunglasses is standing on the street

Próximos Pasos

✓ Utilizar el Dataset **Flickr30k** con más de 3 veces la cantidad de imágenes con las que hemos entrenado.

✓ Probar **nuevas arquitecturas**, diferentes modelos de **word embeddings** (large_model_spacy) y otras **CNNs diferentes a Resnet50 y InceptionV3** para obtener los vectores de las imágenes.

✓ Incorporar **mecanismos** de **atención** para brindarle a red recurrente (decoder) un contexto en donde poner mayor atención y generar mejores predicciones.

Material Bibliográfico

- https://www.kaggle.com/adityajn105/flickr8k **(Dataset)**

- https://academictorrents.com/details/9dea07ba660a722ae1008c4c8afdd303b6f6e53b

- https://machinelearningmastery.com/develop-a-deep-learning-caption-generation-model-in-python/

- https://towardsdatascience.com/image-captioning-with-keras-teaching-computers-to-describe-pictures-c88a46a311b8 **(Referencia: Data Generator Callback y Arquitectura)**

- http://static.googleusercontent.com/media/research.google.com/es//pubs/archive/43274.pdf

Repositorio

¡GRACIAS!