Νευρωνικά Δίκτυα & Ευφυή Υπολογιστικά Συστήματα

Αναφορά 3ης Εργασίας

Ομάδα 99
Αντωνίου Γιώργος 03117715
Κυριάκου Δημήτρης 03117601
Χατζηχριστοφή Χρίστος 03117711

Διαδικασία

1. Ξεκινήσαμε από το tutorial του TensorFlow
2. Κάναμε τις απαραίτητες αλλαγές ώστε να χρησιμοποιεί ως είσοδο το σύνολο δεδομένων flickr30k-images-ecemod
3. Προσθέσαμε τις απαιτούμενες αλλαγές στον κώδικα ώστε να δοκιμαστούν οι προτεινόμενες βελτιώσεις (αλλά και άλλες βελτιώσεις που σκεφτήκαμε) όπως φαίνεται στο μέρος «Βελτιστοποιήσεις»
4. Παραμετροποιήσαμε τον κώδικα και δημιουργήσαμε ένα κελί στην αρχή του notebook από το οποίο γίνονταν οι αλλαγές των υπερπαραμέτρων.
   Παρατήρηση: Αυτό έγινε ώστε να κάνουμε δοκιμές χωρίς να χρειάζεται να αλλάζουμε τον κώδικα σε όλο το notebook αλλά μόνο το πρώτο κελί
5. Κάναμε ανεξάρτητες δοκιμές με διάφορες τιμές για τις υπερπαραμέτρους.
   Παρατήρηση: Λόγω του μεγάλου εύρους πιθανών συνδυασμών η λογική μας ήταν να ξεκινήσουμε με τυχαίες και πολύ διαφορετικές μεταξύ τους δοκιμές και στη συνέχεια να συνδυάσουμε αυτές με τα καλύτερα αποτελέσματα
6. Προσαρμόσαμε το notebook ώστε να λειτουργεί με test set αυτό του διαγωνισμού και να παράγει το απαιτούμενο αρχείο εξόδου
7. Λάβαμε μέρος στο διαγωνισμό με όνομα ομάδας “they see me coding”

Βελτιστοποιήσεις

Προεπεξεργασία κειμένου

Υλοποίηση:

- Για την προεπεξεργασία των captions δημιουργήσαμε υπερπαραμέτρους MIN_CAPTION_LEN και MAX_CAPTION_LEN που περιορίζουν το ελάχιστο και το μέγιστο μήκος caption που χρησιμοποιούνται τόσο για την εκπαίδευση όσο και την αξιολόγηση
- Δημιουργήσαμε την υπερπαράμετρο VOCABULARY_SIZE που καθορίζει το μέγεθος του vocabulary και έχει default τιμή 5000

Παρατήρησεις:

Περιορισμένο εύρος μηκών των captions δίνει καλύτερα αποτελέσματα. Παρατηρήσαμε ότι υπάρχουν κάποια captions στο dataset που έχουν υπερβολικά μεγάλο μήκος και πολλές πληροφορίες που ξεφεύγουν από την ουσία της εικόνας έτσι επιλέξαμε να τις αποκόψουμε. Ακόμη, υπάρχουν κάποια datasets με 2 ή 3 λέξεις που επίσης αποφασίσαμε να τα αποκόψουμε. Κάναμε επίσης δοκιμές αποκόπτοντας και λίγο μεγαλύτερα μήκη (μέχρι 6 λέξεις) αλλά αυτό είχε αρνητική επίδραση στην επίδοση. Δοκιμάσαμε επίσης διάφορες τιμές για το Vocabulary size και φάνηκε ότι οι 5000 λέξεις που είχε αρχικά το tutorial ήταν λίγες.

Encoder

Υλοποίηση:

- Για την επιλογή του έτοιμου μοντέλου CNN που θα χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από τις εικόνες δημιουργήσαμε την υπερπαράμετρο CNN_USED που παίρνει τιμές 'InceptionV3' και 'ResNet152V2', που είναι και το δίκτυο που αντιστοιχεί στον αριθμό της ομάδας μας. Για να γίνει αυτό έγιναν κάποιες τροποποιήσεις γιατί το δίκτυο 'ResNet152V2' δέχεται είσοδο εικόνες με διαστάσεις (224,224) παράγει vectors με διαστάσεις (49, 2048) ενώ το 'InceptionV3' δέχεται είσοδο εικόνες με διαστάσεις (299,299) παράγει vectors με διαστάσεις (64, 2048)

Παρατήρησεις:

Θεωρητικά το δίκτυο ResNet152V2 περιμέναμε ότι θα έχει σταθερά καλύτερα αποτελέσματα και γι' αυτό οι περισσότερες δοκιμές μας χρησιμοποίησαν αυτό. Παρόλα αυτά στην πράξη δεν υπήρχε εμφανής διαφορά και υπήρχαν περιπτώσεις που το InceptionV3 απέδιδε καλύτερα.

Embeddings

Υλοποίηση:

- Για τη χρήση ή όχι έτοιμων embeddings δημιουργήσαμε την υπερπαράμετρο PRETRAINED_EMBEDDINGS που παίρνει τιμές True ή False. Τα έτοιμα embeddings που χρησιμοποιούνται είναι τα glove-wiki που περιέχονται στο φάκελο glove.6B
- Δημιουργήσαμε επίσης την παράμετρο EMBEDDING_DIM που για PRETRAINED_EMBEDDINGS = True καθορίζει τη διάσταση των embedding που θα χρησιμοποιηθούν και παίρνει μία από τις τιμές (50, 100, 200, 300) ενώ για PRETRAINED_EMBEDDINGS = False καθορίζει τη διάσταση των embeddings

Παρατήρησεις:

Η χρήση των έτοιμων embeddings με διάσταση 300 βελτίωσε την επίδοση και τη χρησιμοποιήσαμε αρκετά παρόλο που αυξήθηκε ο χρόνος εκπαίδευσης. Ωστόσο, η χρήση έτοιμων embeddings μικρότερων διαστάσεων δεν είχε το ίδιο θετική επίδραση στην επίδοση. Γενικά η μείωση της διάστασης είχε αρνητική επίδραση στην επίδοση.

Regularization

Υλοποίηση:

- Για τη χρήση ή όχι μηχανισμού ομαλοποίησης με την προσθήκη επιπέδων Dropout μετά από τα δύο πυκνά επίπεδα δημιουργήσαμε τις υπερπαραμέτρους DROPOUT_AFTER_FC1 και DROPOUT_AFTER_FC2 που παίρνουν τιμές True ή False
- Για κάθε dropout layer υπάρχει η αντίστοιχη παράμετρος DROPOUT_VALUE

Παρατήρησεις:

Έγιναν δοκιμές με και χωρίς τα επίπεδα dropout αλλά και με διάφορες τιμές, κοντά στο 0.5, αλλά δεν φάνηκε να επηρεάζεται σημαντική η επίδοση.

Decoder

Υλοποίηση:

- Για την επιλογή του decoder που θα χρησιμοποιείται δημιουργήσαμε την υπερπαράμετρο DECODER_LAYER που παίρνει τιμές ‘GRU’ και ‘LSTM’
- Δημιουργήθηκε επίσης η υπερπαράμετρος UNITS

Παρατήρησεις:

Αν και οι διαφορές στην επίδοση δεν ήταν πολύ εμφανείς, το LSTM φάνηκε να έχει ελαφρώς καλύτερη αποδοση. Επίσης, η μείωση των units είχε αρνητική επίδραση στην επίδοση.

Sentence Generator

Υλοποίηση:

- Για την επιλογή του Sentence Generator που θα χρησιμοποιείται δημιουργήσαμε την υπερπαράμετρο SENTENCE_GENERATOR_METHOD που παίρνει τιμές 'CATEGORICAL' και 'BEAM_SEARCH'. Για την υλοποίηση της Beam Search υλοποιήθηκε συνάρτηση παρόμοια με την evaluate του tutorial που χρησιμοποιεί την tf.random.categorical. Για την υλοποίηση αυτής της συνάρτησης χρησιμοποιήθηκε η πηγή: https://github.com/yashk2810/Image-Captioning/blob/master/Image%20Captioning%20InceptionV3.ipynb
- Εφόσον επιλεχθεί ως DECODER_LAYER το Beam_Search, η υπερπαράμετρος BEAM_WIDTH καθορίζει το πλάτος της ακτίνας που χρησιμοποιεί ο αλγόριθμος δηλαδή το πλήθος καλυτέρων λέξεων που διατηρεί για κάθε βήμα

Παρατήρησεις:

Η χρήση του Beam Search έδινε σταθερά χειρότερα αποτελέσματα για κάθε τιμή b που δοκιμάσαμε. Επιπλέον ήταν σημαντικά πιο αργή από την Categorical. Για αυτούς τους λόγους μετά τις πρώτες δοκιμές σταματήσαμε να τη χρησιμοποιούμε.

Λοιπές Υπερπαραμετροι

Υλοποίηση:

- Δημιουργήθηκε η υπερπαράμετρος EPOCHS που καθορίζει τον αριθμό των εποχών που χρησιμοποιούνται για την εκπαίδευση του μοντέλου

Παρατήρησεις:

Η αύξηση του αριθμού των εποχών μειώνει το Loss και βελτιώνει σημαντικά την επίδοση. Ωστόσο φαίνεται ότι μετά τις 65-70 εποχές φτάνει σε κορεσμό και δεν παράγει πλέον καλύτερα αποτελέσματα. Στα τελικά μας μοντέλα επιλέξαμε τιμές κοντά στο 60 αφού μεγαλύτερες δεν προσέφεραν καλύτερα αποτελέσματα αλλά μόνο αύξαναν τον χρόνο εκπαίδευσης.

Γενικές Παρατηρήσεις:

- Το περιορισμένο λεξιλόγιο εκπαίδευσης δεν επιτρέπει στο μοντέλο να χρησιμοποιήσει σπάνιες, για το dataset, λέξεις που υπάρχουν στα references. Αυτό στοιχίζει στις επιδόσεις με τις μετρικές BLEU. Για παράδειγμα, υπάρχουν εικόνες των οποίων τα references περιέχουν πολύ εξειδικευμένες λέξεις (πχ η παιχνιδοκονσόλα Wii ή ο Νόβακ Τζόκοβιτς).
- Υπάρχουν περιπτώσεις στις οποίες το μοντέλο παράγει καλά captions με αντικείμενα που υπάρχουν στην εικόνα αλλά τα references δεν κάνουν κάποια αναφορά σε αυτά οπότε και πάλι στοιχίζει στις επιδόσεις με τις μετρικές BLEU παρόλο που το αποτέλεσμα είναι καλό.
- Κάθε εκτέλεση της συνάρτησης evaluate παράγει εντελώς καινούργιο caption. Αυτό σημαίνει ότι ακόμα και αν διατηρηθούν σταθερές οι υπερπαραμέτροι ή ακόμα και τα βάρη εκπαίδευσης, για κάθε νέα εκτέλεση οι μετρικές BLEU θα δίνουν διαφορετικό αποτέλεσμα. Συνεπάγεται ότι μία λύση με ελαφρώς καλύτερη επίδοση σε σχέση με κάποια άλλη, σύμφωνα με τον πιο κάτω πίνακα, δεν είναι αρκετό στοιχείο ώστε να θεωρηθεί γενικά καλύτερη λύση.

Δοκιμές

Αναλυτικός κατάλογος δοκιμών

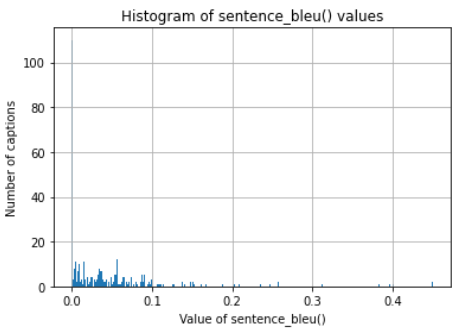
		PREPROCESSING		ENCODER	EMBEDDINGS	REGULARIZATION (DROPOUT)		DECODER	SENTENCE GENERATOR	TRAINING	RESULT
	Images	Caption Lengths	Vocabulary Size	CNN used	Pre-Trained (dim)	after FC1 (value)	after FC2 (value)	Layer (units)	Method	Epochs	Corpus BLEU value
1	6000	[4,30]	10000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (5)	50	0.054477987
2	6000	[7,50]	10000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (5)	50	0.067597591
3	6000	[4,30]	10000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (512)	Categorical	50	0.098663501
4	6000	[4,50]	5000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (5)	50	0.069181942
5	6000	[4,50]	10000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (3)	50	0.09666759
6	6000	[7,50]	10000	ResNet152V2	False (256)	True (0.5)	True (0.5)	LSTM (256)	Categorical	60	0.088078411
7	6000	[7,50]	10000	ResNet152V2	False (200)	True (0.5)	True (0.5)	LSTM (256)	Beam Search (3)	60	0.078690784
8	6000	[3,50]	10000	ResNet152V2	False (256)	False	False	GRU (512)	Beam Search (5)	50	0.054691102
9	6000	[4,50]	10000	ResNet152V2	True (300)	True (0.5)	True (0.5)	GRU (512)	Categorical	60	0.101108075
10	6000	[4,50]	15000	ResNet152V2	True (200)	False	False	GRU (512)	Categorical	60	0.097846439
11	6000	[4,50]	10000	InceptionV3	False (256)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (3)	50	0.077226023
12	6000	[4,50]	10000	InceptionV3	False (256)	True (0.5)	True (0.5)	LSTM (512)	Categorical	50	**0.116282454**
13	6000	[4,50]	10000	ResNet152V2	True (300)	True (0.5)	True (0.5)	GRU (512)	Beam Search (5)	60	0.06823512
14	6000	[4,35]	8000	ResNet152V2	True (300)	True (0.5)	True (0.5)	LSTM (512)	Beam Search (5)	60	0.082260271
15	6000	[4,35]	5000	InceptionV3	True (300)	True (0.45)	True (0.6)	GRU (512)	Categorical	65	0.106058752
16	6000	[4,35]	8000	ResNet152V2	True (300)	True (0.5)	True (0.5)	LSTM (512)	Categorical	60	**0.114765403**

Παρατήρησεις:

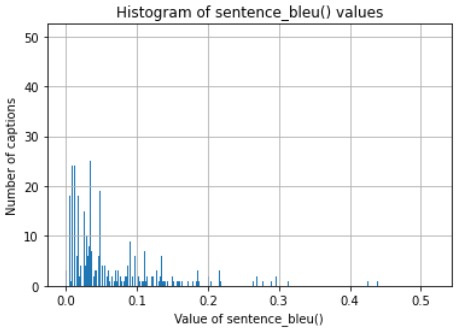
Εκτός από τις πιο πάνω δοκιμές έγιναν και άλλες δοκιμές στα πλαίσια του διαγωνισμού που όμως δεν καταγράφηκαν

Ιστογράμματα επιδόσεων sentence bleu για κάθε δοκιμή

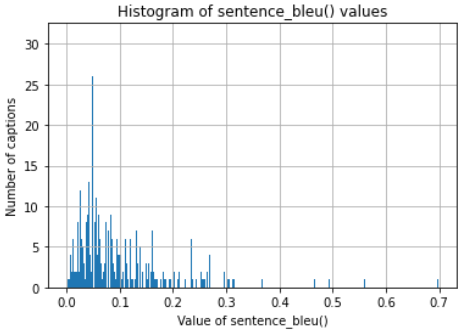
1.



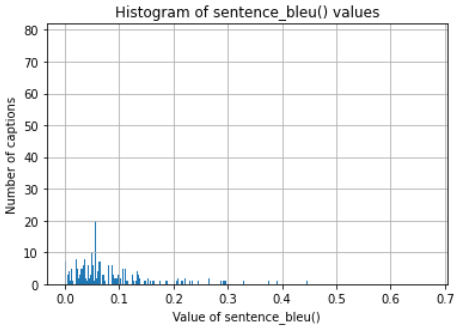
2.



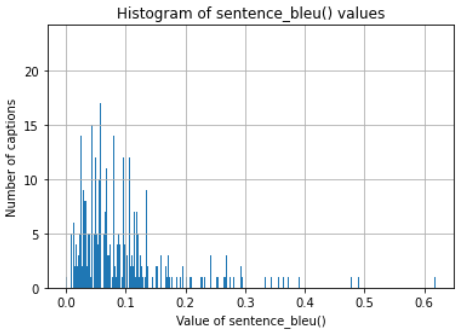
3.



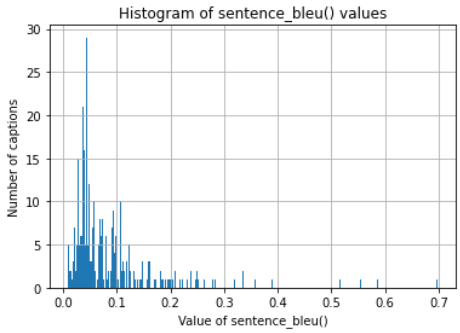
4.



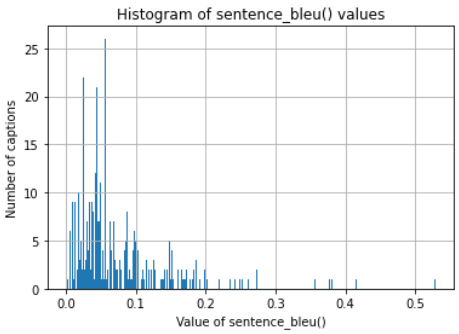
5.



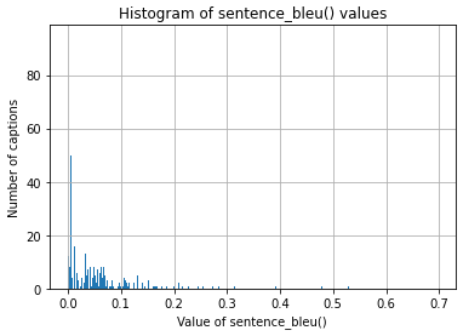
6.



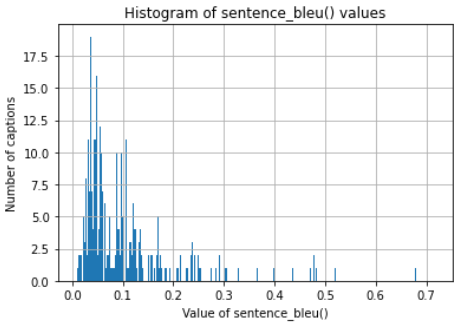
7.



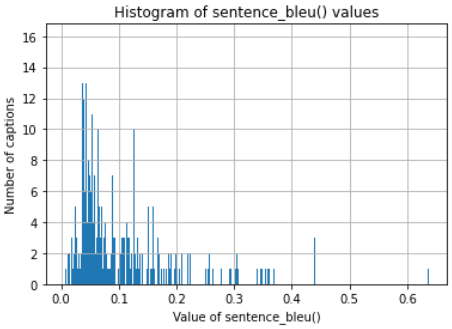
8.



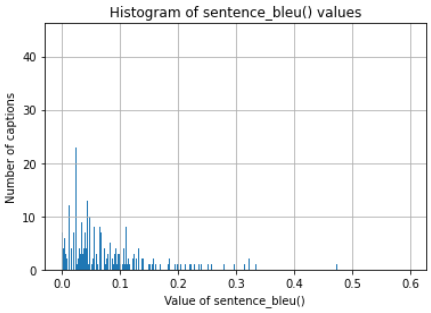
9.



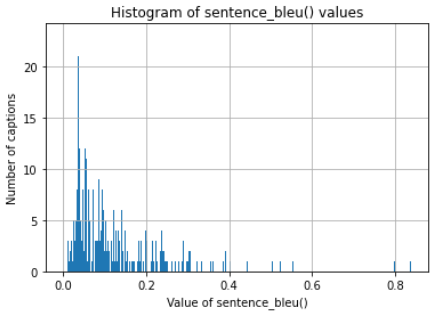
10.



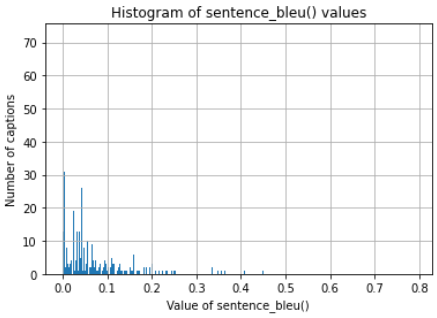
11.



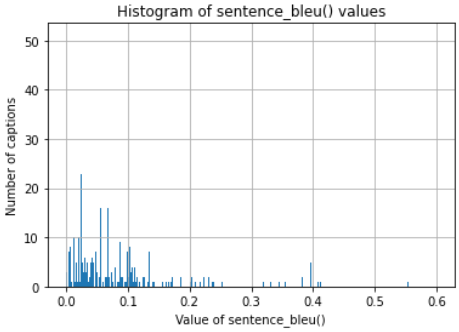
12.



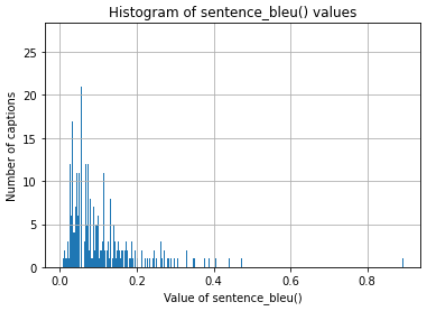
13.



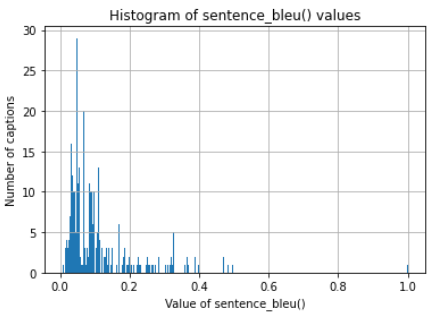
14.



15.

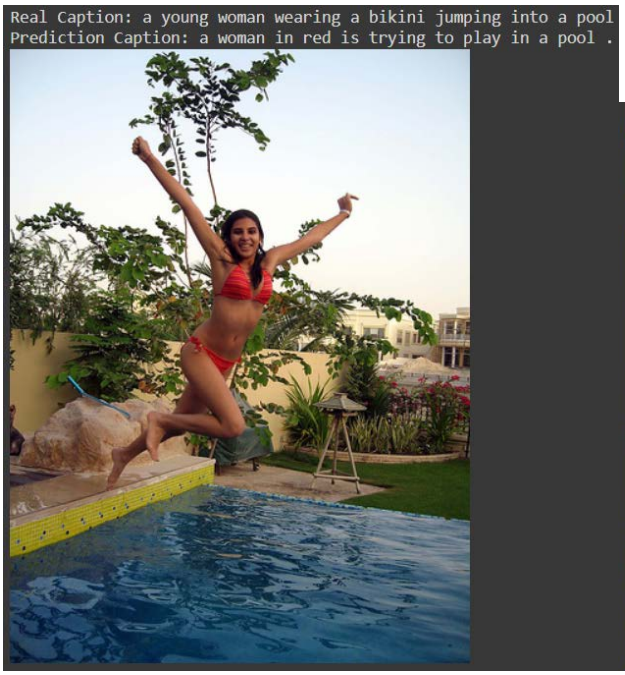


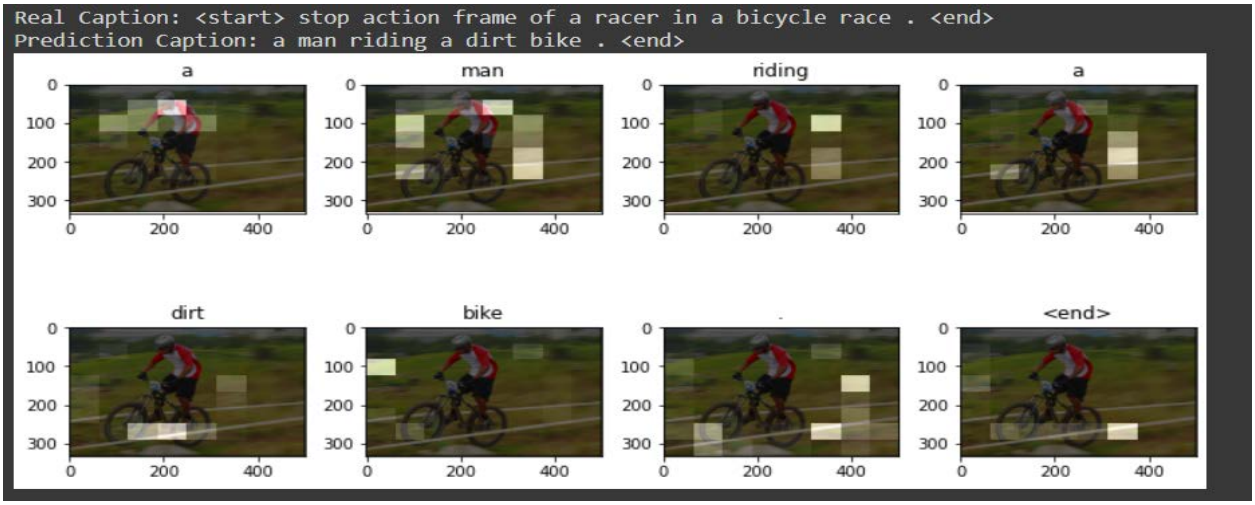
16.

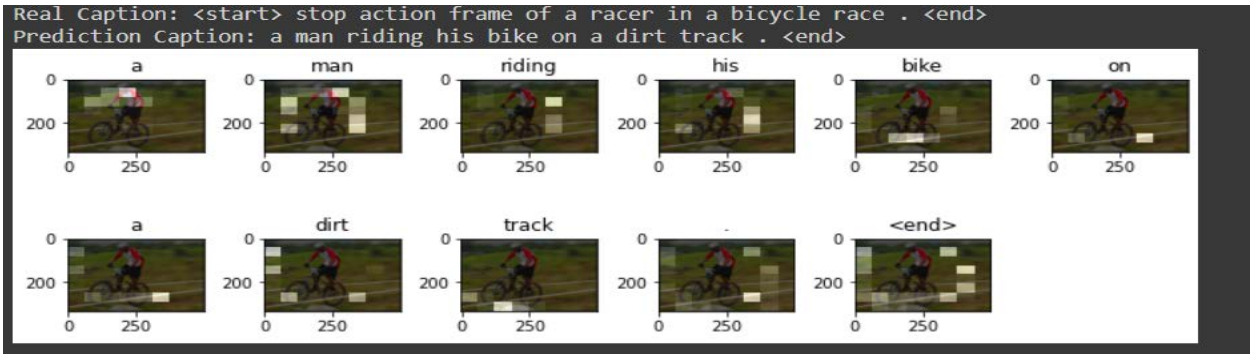


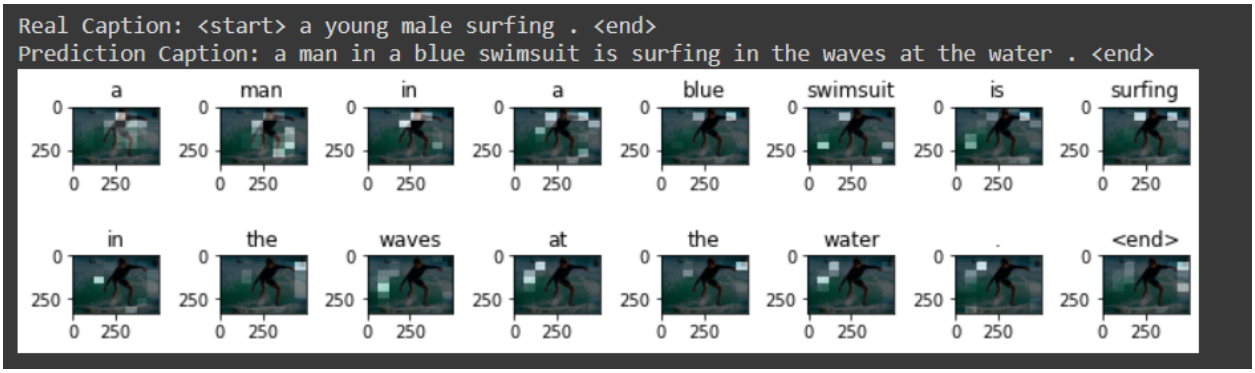
Παραδείγματα εκτελέσεων

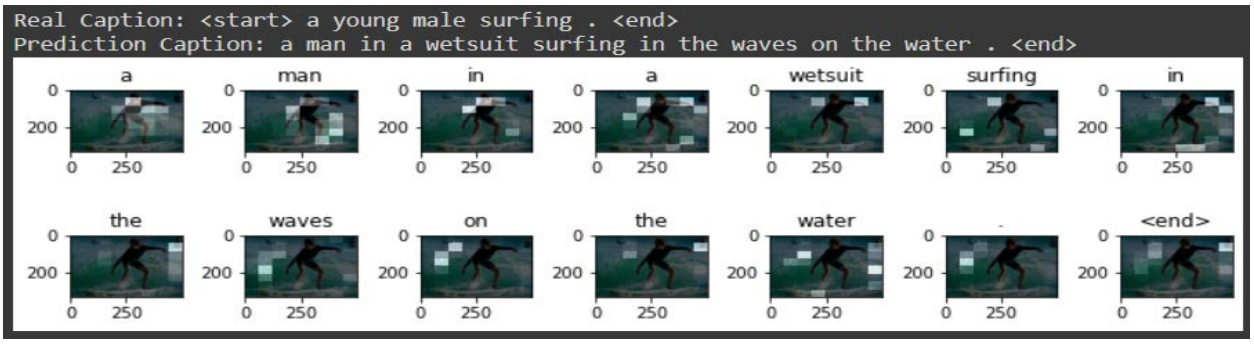
Καλά παραδείγματα




Real Caption: <start> stop action frame of a racer in a bicycle race . <end>
Prediction Caption: a man riding a dirt bike . <end>
a
man
riding
a
dirt
bike
.
<end>


Real Caption: <start> stop action frame of a racer in a bicycle race . <end>
Prediction Caption: a man riding his bike on a dirt track . <end>
a
man
riding
his
bike
on
a
dirt
track
.
<end>

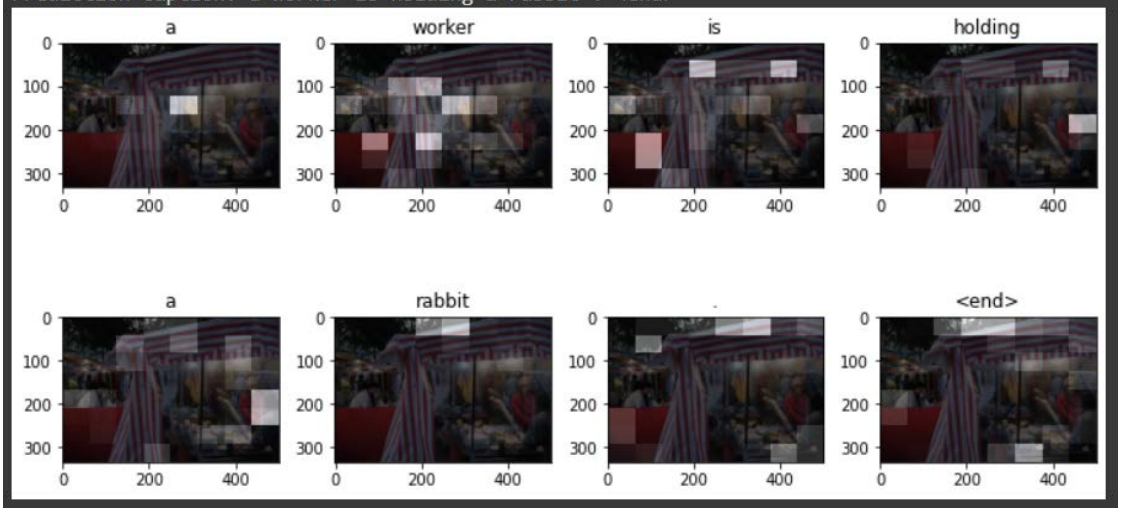

Real Caption: <start> a young male surfing . <end>
Prediction Caption: a man in a blue swimsuit is surfing in the waves at the water . <end>
a
man
in
a
blue
swimsuit
is
surfing
in
the
waves
at
the
water
.
<end>


Real Caption: <start> a young male surfing . <end>
Prediction Caption: a man in a wetsuit surfing in the waves on the water . <end>
a
man
in
a
wetsuit
surfing
in
the
waves
on
the
water
.
<end>

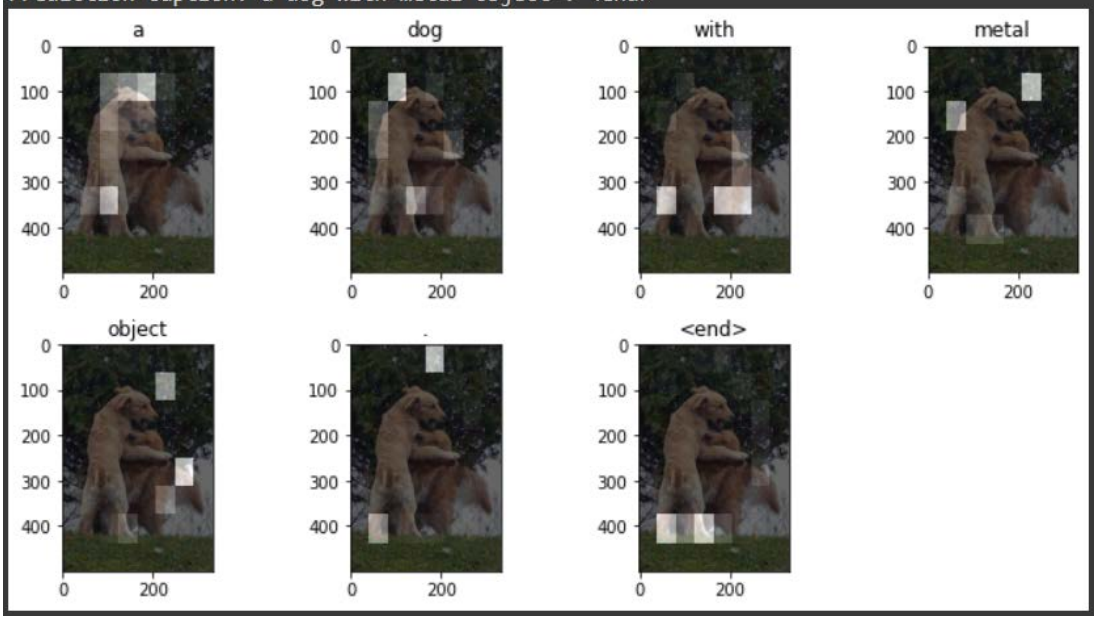
Κακά παραδείγματα

Real Caption: a band performs using medieval instruments at a renaissance fair
Prediction Caption: man and a man in a small clothing dancing

Real Caption: <start> woman in red shirt shopping in a outdoor market . <end>
Prediction Caption: a worker is holding a rabbit . <end>



Real Caption: <start> two tan dogs fight in front of a tree . <end>
Prediction Caption: a dog with metal object . <end>



Real Caption: <start> a girl overlooks a boy sitting down reading a book . <end>
Prediction Caption: a woman in a white shirt stand looking at the computer . <end>