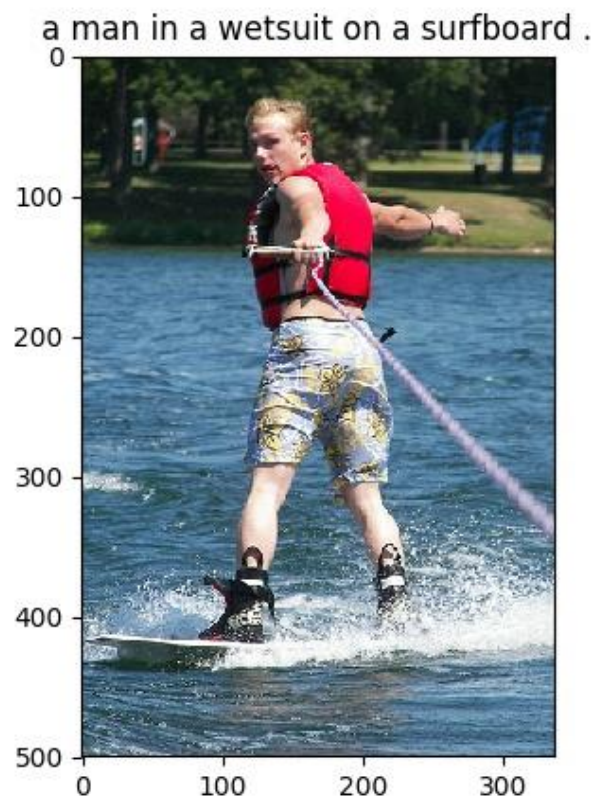


PROJET

NLP & COMPUTER VISION

Prédire la légende d'une image
(Annotation Automatique des images)



Brief

Contexte

La génération de légendes est un problème d'intelligence artificielle difficile où une description textuelle doit être générée pour une photographie donnée.

Il nécessite à la fois des méthodes de vision par ordinateur pour comprendre le contenu de l'image et un modèle de langage du domaine du traitement du langage naturel pour transformer la compréhension de l'image en mots dans le bon ordre. Récemment, les méthodes de Deep Learning ont obtenu des résultats de pointe sur des exemples de ce problème.

L'annotation automatique d'images consiste à associer à chaque image un groupe de mots qui décrit le contenu visuel de l'image au moyen d'un système sans aucune intervention humaine. Cette tâche a fait, et fait toujours, l'objet de nombreux travaux.

L'annotation d'images, dans le cadre de l'apprentissage automatique ou de l'apprentissage profond, est le processus par lequel les images sont étiquetées ou classées à l'aide de texte ou d'outils d'annotation, ou des deux, permettant à un système d'intelligence artificielle (un robot, une machine ou un système automatisé, etc.) de reconnaître par lui-même les caractéristiques des données. Annoter une image, c'est ajouter des métadonnées à un ensemble de données.

La tâche à accomplir est de développer une approche qui utilise un modèle unique de bout en bout pour prédire une légende, à partir d'une photo.

Bien entendu, vous allez devoir gérer ce projet avec une méthodologie précise :

- Découpage des tâches
- Planification
- Présentation du POC
- Rédaction de rapports

Il pourrait être intéressant de commencer à regarder du côté des méthodes « Agile », pour vous aider dans la gestion de projet.

Vous trouverez plus bas les spécifications techniques.

Cahier des charges

- Préparez des données photo et textuelles pour former un modèle de Deep Learning.
 - Flickr8K Dataset
- Concevoir et former un modèle de Deep Learning pour la génération de légendes.
 - CNN avec LSTM ou
 - RNN ...etc
- Évaluez et entraînez le modèle de génération de légendes et utilisez-le pour sous-titrer des photographies entièrement nouvelles.
 - BLEU ou ROUGE
- Interface
 - Maquette
 - Interface Graphique

Spécifications techniques :

NLP & COMPUTER VISION

1. Importer l'ensemble de données de photos et de légendes
 - [Flickr8k Dataset.zip](#)
 - [Flickr8k text.zip](#)
2. Préparer les données photo
 - Reshape les images Selon le modèle choisi
 - Extraction de features
 - Etc.
3. Préparer les données texte
 - Convertir en minuscule
 - Suppression de la ponctuation
 - Suppression des chiffres, etc.
 - Stemming ou Lemmatisation
 - Tokenisation
 - Suppression des mots de fréquence plus élevée (ne contient pas d'informations importantes).
 - Ajoutez deux jetons nommés début et fin dans les légendes
4. Développer un modèle de Deep Learning
 - Transfer Learning
 - VGG16, Inception V3, ResNet ou d'autres
 - Décodeur contenant différentes couches de LSTM
5. Séparez les données d'entraînement et de test
6. Entraînement du modèle à l'aide de données d'entraînement

7. Testez le modèle et évaluez les performances
 - Utilisation des métriques comme BLEU et ROUGE
8. Générer de nouvelles légendes
9. Traduire les légendes en français

Interface

1. Maquette pour les opérateurs
 - Page d'accueil
 - Page de chargement d'une image
 - Page de prédiction d'une légende
2. Réalisation de l'interface
 - Technologie Python / Flask
 - Respecter la maquette définie

1 partie **bonus**

3. Analyse des sentiments :
 - Le texte généré est ensuite soumis à une analyse sentimentale qui génère des sentiments pour un texte particulier.
 - Les sentiments de sortie sont mappés avec des emojis.
 - Fusionnez ensuite les émoticônes avec du texte pour générer une meilleure légende.

Gestion de projet

Vous allez devoir organiser une gestion de projet qui prévoit (avec Trello):

- Des réunions
- Des prises de décisions
- Une planification
- La rédaction d'un rapport
- La préparation d'une soutenance

Attention il sera important de désigner un chef de projet. Il aura pour rôle de s'assurer que les tâches sont bien distribuées et que l'avancement est assuré.

Tips

- Effectuez une réunion de lancement
 - Recontextualiser le projet
 - Prévoyez un plan d'action
 - Découper et planifier vos tâches
- Prévoyez des réunions régulières pour vérifier l'état d'avancement
- Mettez en place un outil pour vérifier l'état d'avancement du projet (Excel, Trello, Gant)

Livrable

Jeudi 04 mai 2023 avant midi

Rapport de projet :

1. Contexte
2. Planification
3. Guide utilisateur
4. Documentation technique
5. Difficultés rencontrées
6. Perspectives d'évolutions
7. Conclusion
8. Bilan de groupe
9. Bilans personnels

Vendredi 05 mai 2023

Jour de la soutenance

Démonstration avec une Interface de prédiction de légendes :

- Il doit répondre aux spécifications techniques
- Être codé avec de bonnes pratiques (code clair et commenté)

Soutenance de Projet

- 20 minutes maximum de présentation avec Slides
 - Intro
 - Planification / Répartition des tâches
 - Démonstration
 - Difficultés rencontrées
 - Perspectives d'évolution
 - Conclusion
 - Bilan de groupe
 - **Bilan perso**
- 10 minutes de question

Ressources

Différentes méthodes :

- [How to Develop a Deep Learning Photo Caption Generator from Scratch \(machinelearningmastery.com\)](https://machinelearningmastery.com/how-to-develop-a-deep-learning-photo-caption-generator-from-scratch/)
 - https://www.renom.jp/notebooks/tutorial/image_processing/neural_image_caption_generator/notebook.html
 - [Develop an image captioning deep learning model using Flickr 8K data \(fairyonice.github.io\)](https://github.com/fairyonice)
-
- Gestion de projet SCRUM :

<https://blog.trello.com/fr/methode-agile-scrum-gestion-projet>

Modalités d'évaluations

Passage par groupe de 4 (ou de 5)

Tous les membres du groupe doivent participer oralement

La soutenance doit être accompagnée d'une présentation visuelle

La soutenance ne doit pas dépasser 20 minutes

La présentation est claire et bien préparée

Les questions ne dépasseront pas 10 minutes

Le jury sera composé du formateur et du responsable pédagogique (il pourra être assisté par une personne extérieure à la formation)

Chaque participant intervient

Le rapport regroupe tous les éléments demandés (il est bien structuré)

La planification montre que chaque membre s'est vu attribuer une tâche

Objectifs visés d'apprentissage

Savoir prétraiter des données non-structurées (texte et images)

Savoir construire des modèles de Deep Learning

Savoir utiliser la technique du Transfer Learning

Savoir mettre en place des réseaux de neurones pour les images et les textes

Savoir réaliser un générateur de légendes

Être capable d'évaluer les performances des modèles avec les métriques appropriées