

Rapport technique de projet de fin d'étude

Lounes ALLIOUI

LOUNES.ALLIOUI@ETU.SORBONNE-UNIVERISTE.FR

Branis GHOUL

BRANIS.GHOUL@ETU.SORBONNE-UNIVERISTE.FR

Amel MEZEMATE

AMEL.MEZEMATE@ETU.SORBONNE-UNIVERISTE.FR

Salomé BOUIN

SALOME.BOUIN@ETU.SORBONNE-UNIVERISTE.FR

Zahra BENSLIMANE

ZAHRA.BENSLIMANE@ETU.SORBONNE-UNIVERISTE.FR

Mehdi KHATI

MEHDI.KHATI@ETU.SORBONNE-UNIVERISTE.FR

*Master Ingénierie des Systemes Intelligents
Sorbonne Université
Paris, France*

Editor: Projet fin d'étude (2022-2023) groupe 4

Abstract

Ce rapport présente, une approche basée sur l'apprentissage profond et un système de vision stéréoscopique pour reconnaître des espèces sous-marines, mesurer leurs dimensions et suivre leurs mouvements. Le modèle a été entraîné sur un jeu de données de plus de 2000 images. Cette étude met en avant le potentiel de fournir des informations précieuses sur l'impact du changement environnemental sur l'évolution des espèces sous-marine. Contribuant, ainsi, aux efforts en matière de conservation et protection de la bio-diversité marine.

Keywords: Intelligence Artificielle, Mask-Rcnn, Dataset, Stéréo vision, Calibration, Computer Vision, espèces sous marines.

The code for this work is available via this link on Github

1. Introduction

Le réchauffement des océans peut entraîner une modification des conditions climatiques qui peuvent être difficiles à supporter pour certaines espèces, ce qui peut entraîner une réduction de la biodiversité sous-marine. La pollution peut également avoir un impact négatif sur les écosystèmes sous-marins en perturbant la qualité de l'eau et en rendant l'environnement hostile pour les espèces qui y vivent. En outre, les activités humaines telles que la pêche intensive et la destruction de l'habitat peuvent également menacer les espèces sous-marines. Il est important de surveiller ces impacts pour garantir la protection et la préservation des écosystèmes sous-marins et de leurs habitants. Donc il est important de suivre ces espèces sous-marines car elles peuvent révéler de nombreuses informations sur l'état de santé de l'environnement marin.

Le CNRS et la fondation de France coordonnés par N. Le Bris Sorbonne Université ont alors proposé un projet qui consistera à reconnaître certaines espèces à partir de leur caractères morphologiques connus, de quantifier l'évolution de leurs densités, paramètres de croissance (taille, forme) et comportement (tracking d'individus), dans le temps. Des méthodes numériques sur utilisées sur des images acquises par des systèmes autonomes en grande profondeur ou en milieu littoral. Nous avons donc eu recours à des technologies avancées telle que l'intelligence artificielle, le machine learning et le traitement d'image. Une interface à été développée pour faciliter l'accès à ces outil aux per-

sonnes de observatoire océanologique de Banyuls-sur-Mer pour leurs recherche. Cette application est aussi proposée aux étudiants du Master Science de la mer, de Licence Sciences de la Nature et aux biologistes.

2. Dataset

2.1 Recherche de données

Les modèles de machine learning basés sur les réseaux de neurones sont des modèles mathématiques complexes qui imitent la structure le fonctionnement du cerveau humain pour apprendre à partir des données (1). Des vastes quantités de données sont généralement utilisés pour l'entraînement de tel structures. Le jeu de données fournit les exemples sur lesquels l'algorithme d'IA peut se baser pour apprendre les relations entre les entrées et les sorties.

Pour former notre algorithme, nous avons créé un jeu de données personnalisé, cela en collectant des données à partir de diverses sources. Nous avons utilisés des méthodes telles que le web scraping, ce qui consiste à développer des scripts pour trouver et automatiquement des images open source sur internet. Nous avons également téléchargé des ensembles de données sur des sites de référence pour les espèces marines, tels que GBIF et Ocean Archive.

L'annotation a été effectuée manuellement en utilisant un outil graphique appelé RoboFlow pour tracer les contours autour des objets ou des régions d'intérêt sur les images. Cet outil permet l'extraction de plusieurs types de fichiers. Nous avons choisi le format COCO (Common Objects in Context), une structure JSON spécifique dictant comment les étiquettes et les métadonnées sont enregistrées pour un jeu de données d'image. Ce format est adapté aux populaires algorithmes open source de détection d'objet et de segmentation.

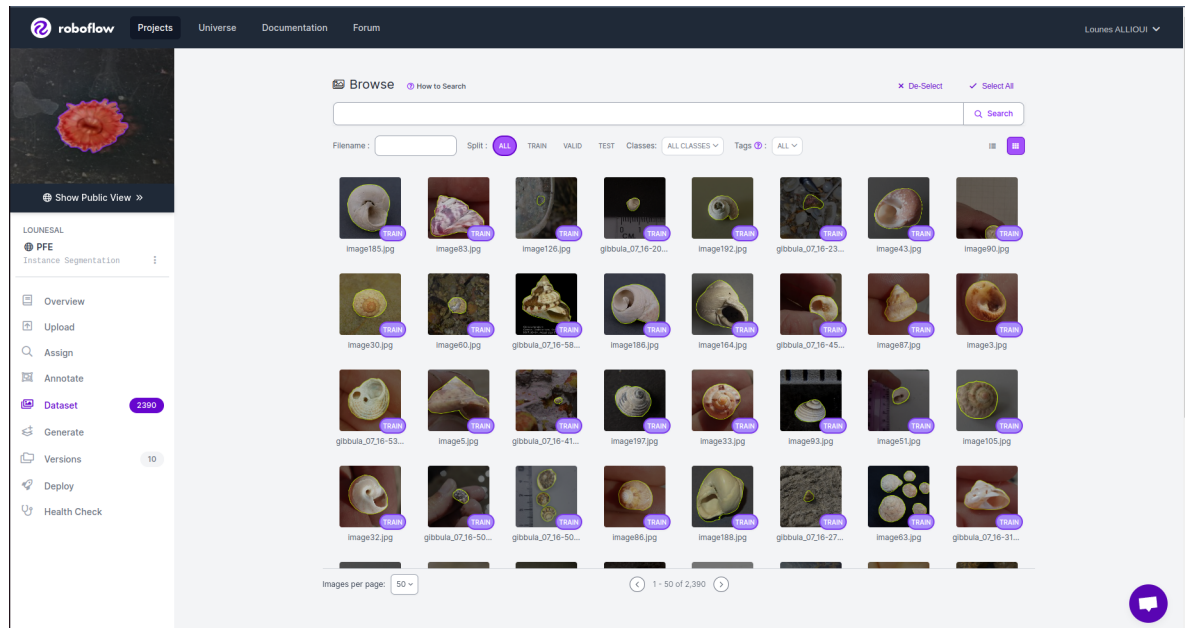


FIGURE 1 – Interface Roboflow pour l'annotation

3. Segmentation

Avant d'extraire les caractéristiques des espèces marines à partir d'images, nous devons d'abord reconnaître et détecter chaque espèce. Cela implique de définir une région dans l'image et d'identifier l'espèce. La reconnaissance d'objets dans des images peut être réalisée avec précision en utilisant l'intelligence artificielle, plus particulièrement la vision par ordinateur. Il existe plusieurs modèles de reconnaissance d'objet basés sur l'apprentissage automatique, tels que les CNN, Fast RCNN, Faster RCNN et Mask RCNN (2) (3) (4) (5). Bien que Fast RCNN, Faster RCNN et les autres soient très puissants et donnent de bons résultats de détection, ils classent les objets en décrivant les régions à l'aide de boîtes englobantes. Nous avons choisi de travailler avec les CNN basées sur Mask RCNN, en particulier, l'algorithme Detectron2 qui a été développé par l'équipe de Facebook AI Research (5), car ce dernier fournit une région plus précise contenant les objets dans l'image en classant chaque pixel. Nous retournons alors tous les pixels contenant l'objet. Ce modèle est plus complexe mais plus précis et, bien que moins rapide dans l'exécution, cela n'affecte pas notre application car nous ne sommes pas intéressés par un traitement en temps réel. La précision est cruciale pour notre application car pour extraire les dimensions des espèces marines, il est important de déterminer avec exactitude les pixels qui les composent.



FIGURE 2 – Résultat de Detectron2

4. Stéréo vision

La mesure de la taille d'un objet est un sujet important dans de nombreuses applications, notamment dans les études de la biologie marine. Cependant, la mesure de la taille d'un objet sous l'eau est un défi en raison des conditions de lumière changeantes, de la visibilité réduite et de la complexité de la capture d'images sous l'eau. Les méthodes courantes utilisent des techniques d'analyse d'image et de photogrammétrie pour mesurer la taille d'un objet, mais ces méthodes nécessitent souvent un arrière-plan de couleur particulière, une architecture de placement de caméra spécifique et des objets à mesurer sur un fond plat.

Il est difficile d'obtenir des résultats précis pour des objets sous-marins en raison de la variabilité de la forme, de la taille, de la couleur et de l'orientation de ces objets. Ces méthodes peuvent ne pas être adaptées pour notre application, car elles ne reflètent pas les conditions réelles de

l'environnement sous-marin. Il est donc nécessaire de réfléchir à une méthode plus puissante qui nécessite un matériel peu coûteux. Pour cela nous avons choisi de travailler avec la technique de la triangulation basée sur la stéréo vision qui consiste à utiliser deux images prises à partir de positions différentes pour calculer la distance à l'objet.

Cette méthode présente plusieurs avantages par rapport aux autres méthodes, telles que la précision, la robustesse aux conditions de lumière changeantes, la moindre dépendance de l'arrière-plan et la moindre dépendance de l'orientation de l'objet. En utilisant des informations de distance pour calculer la taille de l'objet, la technique de triangulation permet d'obtenir des résultats plus précis, moins sensibles aux variations de luminosité et moins dépendants de l'arrière-plan et de l'orientation de l'objet. Cette méthode est particulièrement prometteuse pour les études de la biologie marine, car elle permet de mesurer la taille des objets sous-marins dans des conditions réelles. En outre, elle nécessite moins de matériel coûteux par rapport aux autres techniques.



FIGURE 3 – Camera stereo omega

La vision stéréo est une technique permettant de déterminer la structure 3D d'une scène à partir de deux ou plusieurs images 2D prises à partir de points de vue différents. Cela fonctionne en capturant des images d'une scène à l'aide de deux caméras différentes, appelées caméra gauche et caméra droite, qui sont généralement séparées par une faible distance. En comparant les pixels correspondants dans les images gauche et droite, le système peut déterminer la profondeur de chaque pixel dans la scène, ce qui permet de calculer la position 3D des objets dans la scène.

4.1 Trouver la position d'un pixel dans l'espace 3D :

Le principe de base derrière la vision stéréo est celui de la triangulation [cité]. La triangulation est le processus de détermination de l'emplacement d'un point dans l'espace 3D (position réelle en mètres) en connaissant ses emplacements connus dans deux ou plusieurs images (distances en pixels). Dans le cas de la vision stéréo, le point est un pixel dans les images gauche et droite, et les emplacements connus sont les positions des caméras gauche et droite. En résolvant l'intersection des rayons qui projettent des caméras vers le pixel en question, la position 3D du pixel peut être déterminée.

Une fois que la carte de profondeur de l'image stéréo est obtenue, les dimensions des objets peuvent être obtenues à travers des calculs mathématiques simples tels que la triangulation et la

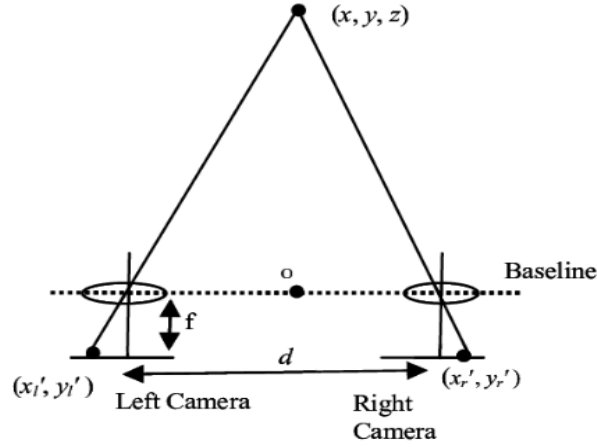


FIGURE 4 – Principe de la stéréo vision

transformation de perspective.

Pour pouvoir calculer les dimensions des objets avec précision, nous avons besoin de connaître les paramètres de la stéréo calibration des caméras gauche et droite (telles que la distance focale et la matrice de projection) ainsi que les positions relatives des caméras par rapport à l'objet que nous voulons mesurer.

4.2 Stéréo calibration des deux caméras

La calibration stéréo des caméras est le processus de détermination des paramètres de caméra pour deux caméras (gauche et droite) utilisées pour la vision stéréo. Ces paramètres comprennent la matrice de projection, la distorsion des lentilles, et les relations de position et de rotation entre les deux caméras. La calibration stéréo des caméras est importante car elle permet de garantir que les images capturées par les caméras gauche et droite sont correctement alignées avant d'être utilisées pour calculer la profondeur. Elle permet également d'éliminer les distorsions de l'image causées par les lentilles, ce qui améliore la qualité des images et la précision des calculs de profondeur. Il existe plusieurs méthodes de calibration stéréo des caméras, mais l'une des méthodes les plus courantes est appelée la méthode de la grille d'échiquier. Il consiste à placer une grille d'échiquier à différents angles et positions devant les caméras et à utiliser les images capturées pour déterminer les paramètres de caméra.

4.3 Calcul des points 3d et les dimensions des espèces

La vision stéréo peut être utilisée pour calculer la longueur et la largeur des espèces sous-marines en utilisant les méthodes de triangulation et de transformation de perspective que j'ai mentionnées précédemment en utilisant les paramètres de calibration de caméras obtenus par la calibration stéréo. Pour obtenir des résultats précis, il est nécessaire de prendre des images de haute qualité et d'utiliser des méthodes de traitement d'image avancées pour éliminer les effets de l'eau et des ombres sur les images. Il peut également être nécessaire de prendre des images à différents angles et de combiner les informations pour obtenir des résultats précis. Dans notre cas, nous pouvons utiliser deux caméras

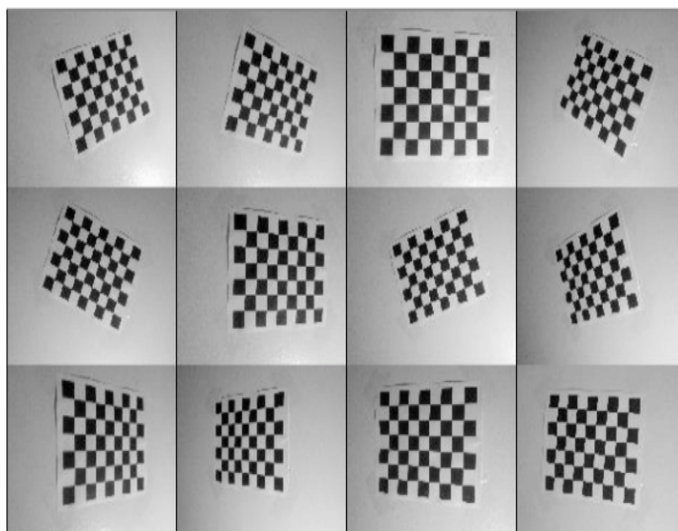


FIGURE 5 – mire de calibration

GoPro fixées sur un axe avec une option permettant de modifier la distance entre les caméras pour trouver la distance idéale. Il est également nécessaire de prendre les deux images en même temps.

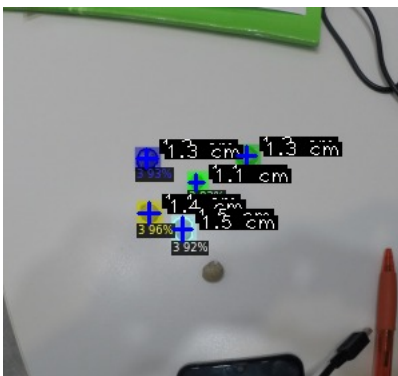


FIGURE 6 – Calcul de dimensions (Image temporaire)

5. Tracking

Le suivi des espèces marines permet de comprendre leur évolution dans le temps et dans l'espace et peut aider à surveiller les impacts environnementaux sur leur santé et leur populations. Cela peut également améliorer notre compréhension des interactions entre différentes espèces et écosystèmes marins et faciliter la prise de décisions pour la gestion de la faune et de la flore marines et la protection des espèces en danger. Nous avons développé un algorithme de suivi des changements de couleur des espèces marines dans le temps en utilisant deux méthodes que nous avons jugées efficaces.

La première méthode utilise les filtres de Kalman, qui permettent de créer des matrices de l'état (F), de mesure (H) et d'incertitude (Q, R) en fonction du système de détection de couleur. Par exemple, pour suivre les changements de couleur de l'espèce, nous utilisons une caméra qui fournit

des matrices à 2 dimensions (x, y) pour suivre sa position et à 3 dimensions (R, G, B) pour suivre les niveaux de chaque canal de couleur. Nous initialisons les valeurs d'état et d'incertitude en utilisant des mesures ou des estimations initiales ou en utilisant directement les premières images capturées par la caméra. Ensuite, nous utilisons une boucle pour itérer sur chaque image du timelapse et les fonctions de mise à jour de Kalman pour mettre à jour l'état et l'incertitude en fonction des nouvelles mesures de couleur. Finalement, pour détecter les changements significatifs dans les niveaux de couleur de l'espèce, nous utilisons les valeurs d'état. Ainsi, le filtre de Kalman est utilisé pour le tracking de couleur.

La deuxième approche consiste à utiliser des techniques de traitement d'image. Nous récupérons d'abord les pixels de l'objet d'intérêt dans chaque image et mesurons les changements de couleur de l'espèce au fil du temps. Nous extrayons d'abord les intensités de chaque pixel à partir de la boîte englobante, puis nous calculons la couleur moyenne. Nous répétons ce processus pour l'image suivante afin de comparer les résultats. Si la différence est supérieure ou inférieure à un certain pourcentage, nous considérons qu'il y a eu un changement de couleur de l'espèce, sinon, nous considérons qu'il s'agit d'un changement de lumière, de bruit, etc.

6. Utilisation de l'interface

L'interface graphique développée pour ce projet est une composante clé qui permet d'interagir avec les données et les résultats obtenus. Conçue en utilisant le langage Python avec la librairie PySide2, elle offre une interface intuitive et facile à utiliser pour les utilisateurs non familiarisés avec les codes informatiques. Cette interface graphique est essentielle pour visualiser les données, explorer les résultats et faire des analyses sur les espèces sous-marines et leur impact sur l'environnement marin.

6.1 Le bouton Calibration

La calibration est une étape importante pour assurer un bon fonctionnement du système de caméra. Elle permet d'ajuster les deux caméras pour un environnement donné. Pour ce faire, il suffit de charger les images d'une mire prises par les deux caméras dans les dossiers correspondants, puis de cliquer sur le bouton **Calibrer**. Il est recommandé de prendre environ 40 images par caméra pour obtenir une précision optimale, avec une erreur inférieure à 0.5.

6.2 Tracking

Cette section permet de suivre les couleurs et de compter le nombre d'individus d'une espèce donnée. Pour cela, il suffit de sélectionner une vidéo provenant d'une seule caméra, puis de cliquer sur le bouton **Commencer**. Les résultats seront affichés à l'écran.

6.3 Image segmentation

Cette partie consiste à calculer les dimensions d'une ou plusieurs espèces à partir de deux images, une de chaque caméra. Après le calcul, les résultats seront immédiatement affichés à l'écran.

6.4 Vidéo segmentation

Cette section est similaire à la précédente, mais cette fois deux vidéos sont utilisées, une pour chaque caméra. Le but est d'afficher les dimensions de chaque espèce détectée dans les vidéos.

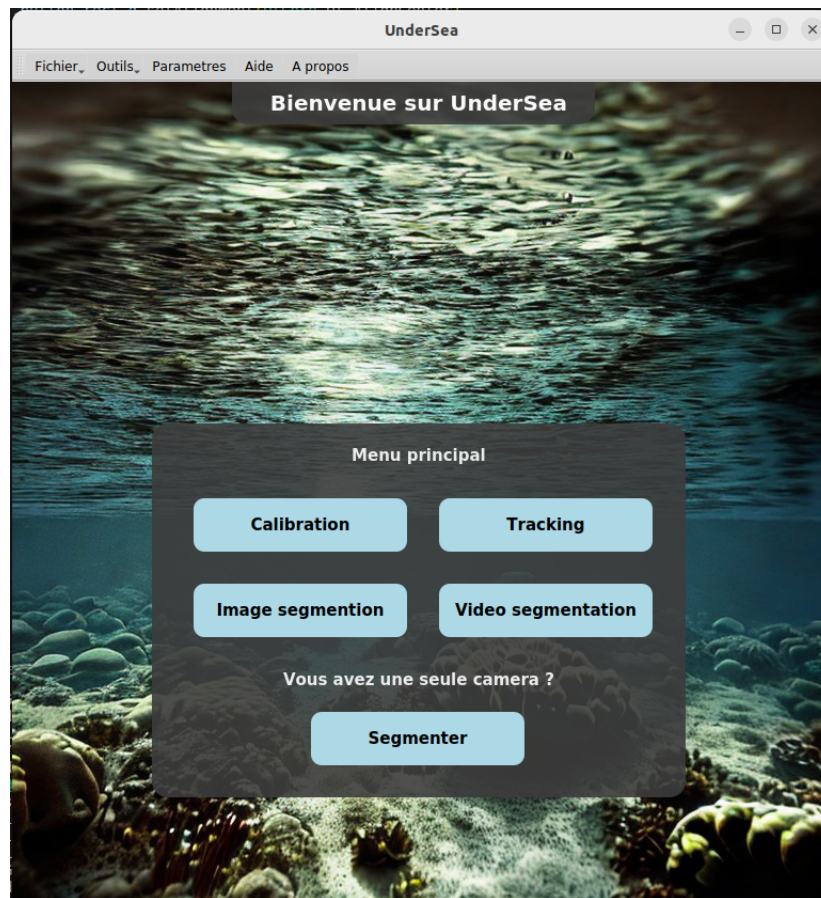


FIGURE 7 – Calcul de dimensions (Image temporaire)

6.5 Segmenter

Cette section est conçue pour les cas où un seul système de caméra est utilisé. Elle permet d'entrer soit une image, soit une vidéo, et d'afficher les résultats de détection et de segmentation pour chaque espèce.

6.6 Barre du menu

Dans les paramètres on peut changer le chemin du modèle d'intelligence artificielle. il est recommandé de ne pas le modifier.

7. Conclusion

L'utilisation de l'apprentissage profond pour résoudre ce problème complexe a montré des résultats prometteurs, mais il est également important de souligner que le modèle peut être affecté par des différences entre les données d'entraînement et de test.

La méthode de segmentation présentée dans ce rapport a été testée et approuvée sur des données similaires à celles utilisées pour l'entraînement des algorithmes. Cependant, Les résultats de ce modèle peuvent chuter si les données testées ne correspondent pas à la même distribution des données d'entraînement. Pour améliorer les performances du modèle dans cette situation, il est important d'utiliser un jeu de données d'entraînement diversifié et représentatif des données cibles que vous souhaitez analyser après l'apprentissage.

Pour finir, ce projet a été un défi technique important. Mais, nous avons appris énormément de choses au cours de son développement. Nous avons compris que les données sont un facteur clé dans les applications de la vie réelle et c'est d'autant plus compliqué pour des applications

Références

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [2] X. Ji, Y. Lu, and L. Guo, “Image Super-Resolution with Deep Convolutional Neural Network,” in *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*. Changsha, China : IEEE, Jun. 2016, pp. 626–630. [Online]. Available : <http://ieeexplore.ieee.org/document/7866198/>
- [3] R. Girshick, “Fast R-CNN,” in *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile : IEEE, Dec. 2015, pp. 1440–1448. [Online]. Available : <http://ieeexplore.ieee.org/document/7410526/>
- [4] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. [Online]. Available : <http://ieeexplore.ieee.org/document/7485869/>
- [5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” 2017. [Online]. Available : <https://arxiv.org/abs/1703.06870>