الجزائرية الديمقراطية الشعبية الجمهورية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 – بسيدي بلعباس

Ecole Supérieure en Informatique

-08 Mai 1945- Sidi Bel Abbes

Mémoire de Fin d'étude

Pour l'obtention du diplôme d'**ingénieur d'état**

Filière : **Informatique**

Spécialité : **Ingénierie des Systèmes Informatiques (ISI)**

Thème

Implementation of A Temporal

Video Coder Using Deep Neural Networks

Présenté par **:**

- Mlle Soumia Zohra EL MESTARI

Soutenu le : **24/09/2020**

Devant le jury composé de :

- Mme BELALIA Amina	Président
- M CHAIB Souleyman	Examinateur
- M BENSLIMANE Sidi Mohamed	Encadreur
- Mme MOKRAOUI Anissa	Co-Encadreur

Année Universitaire : 2019 / 2020

Abstract

During the last few years, the image and Video Compression technologies have grown by leaps and bounds. However, due to the popularization of image and video acquisition devices,the growth rate of image and video data is far beyond the improvement of the compression ratio. Moreover, it has been widely recognised that there are increasing challenges of pursuing further coding performance improvement within the traditional hybrid coding framework.

Currently, existing standards perform poorly with specific contents. Thus, we find a tendency towards applying Neural Networks precisely and data driven techniques in general for Content based Compression.

Capturing temporal redundancy for Video Coding purposes using Deep Learning can be done either Explicitly or Implicitly.

In an explicit manner, the system has to be designed around a neural network unit made only to estimate motion given two successive frames. Such a system require another sub system to perform frame reconstruction given a previous frame and motion informations.This method is complex where the performance of the whole system rely entirely on the ability of the Motion Estimation unit.

From another angle, in an implicit manner,a video can be seen as a set of frames related to one-another by a conditional temporal distribution. Hence by capturing and estimating this distribution using Variational Auto-Encoders, the video can be successfully projected into a lower dimensional space.

This work have investigated both approaches, where the latest End-To-End Video Compression Using VAEs outperformed the latter one using Explicit Motion Estimation. Several design choices and techniques have been applied which led to competitive results. The variational Auto-encoder achieved a good reconstruction quality, PSNR reached 29.02 for related content video clips. While with explicit motion estimation the reconstruction scored lower: PSNR of 20.06 for video clips with motion patterns that similar to the the Flying Chairs dataset motion patterns.

Keywords: Image processing, Video Compression, Deep Neural Networks, Generative Models, Motion Estimation, Convolutional Neural Networks(CNN), Variational Auto-Encoder(VAE).

Résumé:

Au cours des dernières années, les technologies de compression d'images et de vidéo ont connu une croissance fulgurante. Cependant, en raison de la popularisation des dispositifs d'acquisition d'images et de vidéos, le taux de croissance des données d'images et de vidéos est bien supérieur à l'amélioration du taux de compression. En outre, il est largement reconnu que la poursuite de l'amélioration des performances de codage dans le cadre du codage hybride traditionnel pose de plus en plus de problèmes.

Actuellement, les normes existantes sont peu performantes pour des contenus spécifiques. Ainsi, nous constatons une tendance à appliquer les réseaux neuronaux principalement, et les techniques axées sur les données en général pour la compression basée sur le contenu.

La capture de la redondance temporelle pour le codage vidéo à l'aide de l'apprentissage approfondi peut être effectuée de manière explicite ou implicite.

De manière explicite, le système doit être conçu autour d'une unité de réseau neuronal faite uniquement pour estimer le mouvement à partir de deux images successives. Un tel système nécessite un autre sous-système pour effectuer la reconstruction de l'image à partir d'une image précédente et des informations de mouvement. Une telle méthode est complexe lorsque la performance de l'ensemble du système dépend entièrement de la capacité de l'unité d'estimation du mouvement.

Sous un autre angle, de manière implicite, une vidéo peut être vue comme un ensemble d'images liées les unes aux autres par une distribution temporelle conditionnelle. Par conséquent, en capturant et en estimant cette distribution à l'aide d'auto-encodeurs variationnels, la vidéo peut être projetée avec succès dans un espace dimensionnel inférieur.

Ce travail a étudié les deux approches, où la compression vidéo de bout en bout utilisant les VAE a surpassé la dernière en utilisant l'estimation de mouvement explicite.Plusieurs modifications au niveau des architectures neuronales ont permis

d'obtenir des résultats compétitifs. L'auto-codeur variationnel a atteint une bonne qualité de reconstruction , PSNR a atteint 29,02 pour les clips vidéo de contenu connexe. Alors qu'avec l'estimation explicite du mouvement, la reconstruction a obtenu un score inférieur : PSNR de 20,06 pour les clips vidéo dont les modèles de mouvement sont similaires à ceux de l'ensemble de données Flying Chairs.

Mot Clés: Traitement des Images, Compression des Videos, Réseaux Neuronaux Profonds, Modèles Génératifs, Estimation de Mouvement, Réseaux de Neuronnes Convolutionnels(CNN), Auto-Encodeur Variationel (VAE).