

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 - بسبدي بلعباس
Ecole Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbas



MEMOIRE

En vue de l'obtention du diplôme de **Master**

Filière : **Informatique**

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

Thème

Temporal Video Compression

Using Neural Networks

Présenté par :

- Mlle Soumia Zohra EL MESTARI

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

Devant le jury composé de :

- | | | |
|-----------------------------|-------------------------|--------------|
| - Mme BELALIA Amina | Maître de conférences B | Président |
| - M CHAIB Souleyman | Maître de conférences B | Examineur |
| - M BENSLIMANE Sidi Mohamed | Professeur | Encadreur |
| - Mme MOKRAOUI Anissa | Professeur | 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, 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.

The already existing temporal coders, rely entirely on non-dense motion estimation. Despite their performance, such methods fail with specific video contents. Thus the need to more robust and content sensitive coders using Data Driven techniques.

Over the past two decades, the technology lived its greatest breakthrough over multiple fields due usage of Deep Learning for several tasks. Imagery and in particular video compression was no exception.

In this thesis we presented a state of the art about Capturing temporal redundancy for Video Coding purposes using Deep Learning.

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. Such a method is complex where the performance of the whole system rely entirely on the ability of the motion estimation unit in capturing the temporal dependency.

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.

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, le taux de croissance des données d'image et de vidéo est bien au-delà de l'amélioration du taux de compression. En outre, la poursuite de l'amélioration des performances dans le cadre du codage hybride traditionnel constitue un défi de plus en plus important.

Les méthodes existantes de codage temporel, reposent entièrement sur l'estimation non-dense de mouvement. Malgré leurs performances, ces méthodes sont limitées avec des contenus vidéo spécifiques. D'où la nécessité d'avoir de codeurs plus robustes et plus sensibles au contenu, en utilisant des techniques basées sur les données.

Au cours des deux dernières décennies, la technologie a connu sa plus grande percée dans de nombreux domaines grâce à l'apprentissage approfondi. L'imagerie et la compression vidéo n'ont pas fait exception à la règle.

Dans cette thèse, nous avons présenté un état de l'art sur la capture de la redondance temporelle à des fins de codage vidéo à l'aide de l'apprentissage approfondi.

Dans une première manière dite "explicite", le système est 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 à capturer la dépendance temporelle.

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. Ainsi, 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.

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

الملخص

خلال السنوات القليلة الماضية ، نمت تقنيات ضغط الصور والفيديو على قدم وساق. ومع ذلك ، فإن معدل نمو بيانات الصور والفيديو يتجاوز بكثير تحسين نسبة الضغط. علاوة على ذلك ، تم الاعتراف على نطاق واسع بأن هناك تحديات متزايدة لمواصلة تحسين أداء الترميز في إطار الترميز الهجين التقليدي.

حاليًا ، المعايير المتواجدة تعتمد كليًا على تقدير الحركة غير الكثيفة. على الرغم من أدائها ، تقشل هذه الأساليب مع محتويات فيديو محددة. وبالتالي الحاجة إلى برامج تشفير أكثر قوة حساسية للمحتوى باستخدام تقنيات تستند إلى البيانات.

على مدى العقدين الماضيين ، عاشت التكنولوجيا أكبر قفزة لها في مجالات متعددة بسبب استخدام التعلم العميق لعدة مهام. لم تكن الصور وخاصة ضغط الفيديو استثناء.

في هذه الأطروحة استعرضنا الدراسات السابقة حول النقاط التكرار الزمني لأغراض ترميز الفيديو باستخدام التعلم العميق.

بطريقة صريحة ، يجب تصميم النظام حول وحدة شبكة عصبية مصنوعة فقط لتقدير الحركة نظرا لإطارين متتاليين. يتطلب مثل هذا النظام نظاما فرعا آخر لإجراء إعادة بناء الإطار نظرا للمعلومات الإطار والحركة السابقة. هذه الطريقة معقدة حيث يعتمد أداء النظام بأكمله بالكامل على قدرة وحدة تقدير الحركة في النقاط التبعية الزمنية.

من زاوية أخرى ، بطريقة ضمنية ، يمكن رؤية الفيديو كمجموعة من الإطارات المتعلقة ببعضها البعض من خلال توزيع زمني مشروط. وبالتالي من خلال التقاط وتقدير هذا التوزيع باستخدام الترميز التلقائي المتغير ، يمكن عرض الفيديو بنجاح في مساحة أبعاد أقل.

الكلمات الرئيسية: معالجة الصور ، ضغط الفيديو ، الشبكات العصبية العميقة ، النماذج التوليدية ، تقدير الحركة ، الشبكات العصبية التلافيفية (CNN) ، التشفير التلقائي المتغير (VAE).