

Détection d'anomalies en temps réel dans le flux vidéo

Fabien Poirier*, Rakia Jaziri** Camille Srour*** Gilles Bernard****

*fab_16@hotmail.fr, **rjaziri@ai.univ-paris8.fr, ***cs@othello.group,
****gb@ai.univ-paris8.fr,

Résumé. De nos jours, de nombreux lieux profitent de la télésurveillance. Mais lorsqu'un incident survient, celle-ci est utilisée dans le but de constater les événements passés. On peut donc la considérer comme un outil de dissuasion plutôt que de détection. Dans cet article, nous allons proposer une approche d'apprentissage automatique profond (deep learning en anglais) visant à résoudre cette lacune. Cette approche utilise des modèles de convolution (CNN) permettant d'extraire des caractéristiques pertinentes liées aux images analysées, qui formeront par la suite des séries temporelles destinées à être traitées par des modèles récurrents.

1 Introduction

Actuellement, la télésurveillance permet de contrôler à distance plusieurs lieux simultanément grâce à un système de caméras disposé dans un espace public ou privé. Les images obtenues par les caméras sont transmises sur un ensemble d'écrans pour être visionnées et analysées, puis archivées ou détruites. Cette surveillance a pour but de contrôler les conditions de sécurité et de sûreté de ces lieux. Généralement, ces images sont analysées par des personnes physiques, ce qui rend la tâche longue et coûteuse. De plus, l'efficacité d'un tel système dépend de l'attention et de la réactivité du surveillant. Avec l'avancée de l'intelligence artificielle dans de nombreux domaines comme celui du traitement d'image, du traitement audio, de la reconnaissance d'actions, ou encore de la détection d'anomalies, il serait plus efficace d'automatiser au moins partiellement cette analyse dans le but d'aider le surveillant dans sa tâche voire de le remplacer. Pour contribuer à cette tâche, tout modèle se doit d'être capable de détecter d'éventuels problèmes le plus rapidement possible avec un maximum de précision, voir d'être capable de les prédire. Nous abordons dans ce qui suit la question de savoir s'il est possible de réaliser un modèle de détection, en temps réel, d'anomalies dans le flux vidéo.

Dans cet article, nous proposons une approche profonde basée sur les modèles récurrents. De par sa nature complexe, une vidéo peut être analysée de trois manières différentes, suivant qu'on prend en compte l'image, le son, ou les deux à la fois. Étant donné que la majorité des vidéos provenant de caméra de surveillance ne contiennent pas de son, ce qui optimise leur stockage, nous avons focalisé nos travaux sur l'analyse d'image. De plus, nous nous concentrerons principalement sur des anomalies pouvant avoir un impact direct sur la sécurité ou la sûreté des personnes présentes dans les vidéos analysées. Après un bref état de l'art, nous présentons notre architecture, nos expérimentations et résultats, avant de conclure.