



HAL
open science

Optimisation du décodage par liste de vidéos corrompues basée sur une architecture d'apprentissage en profondeur

Yujing Zhang, Stéphane Coulombe, François-Xavier Coudoux, Patrick Corlay

► **To cite this version:**

Yujing Zhang, Stéphane Coulombe, François-Xavier Coudoux, Patrick Corlay. Optimisation du décodage par liste de vidéos corrompues basée sur une architecture d'apprentissage en profondeur. Mardi des chercheurs, Jun 2023, Valenciennes (Nord), France. 2023. hal-04139241

HAL Id: hal-04139241

<https://hal.univ-lille.fr/hal-04139241v1>

Submitted on 23 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

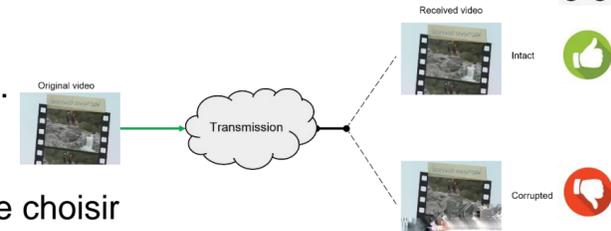
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimisation du décodage par liste de vidéos corrompues basée sur une architecture d'apprentissage en profondeur

Yujing ZHANG, Stéphane Coulombe, François-Xavier Coudoux, Patrick Corlay

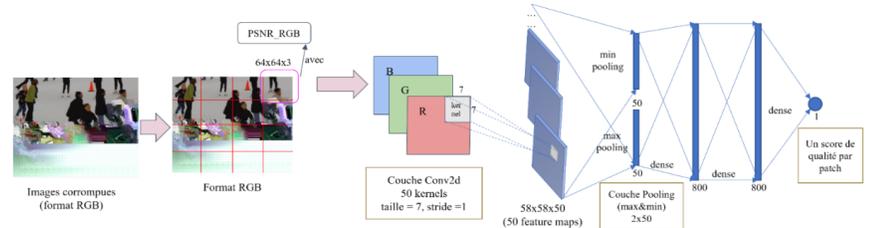
Contexte

- Développement extrêmement rapide des applications impliquant la transmission de contenus vidéo.
- Des erreurs de transmission importantes sur la qualité visuelle des contenus vidéos reçus.
- Utilisation d'une approche de décodage par liste [1] qui exploite les paquets reçus corrompus.
- Génération de plusieurs paquets candidats à partir de chaque paquet corrompu.
- Estimation sans référence au récepteur de la qualité de chacun de ces candidats pour ensuite choisir le meilleur qui correspondra idéalement à la version intacte originalement transmise.

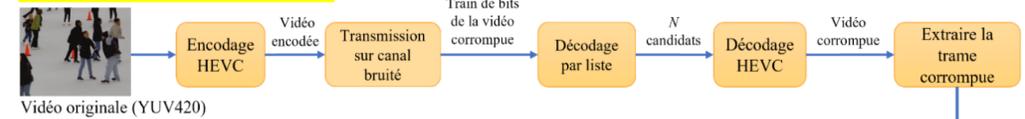


La méthode proposée

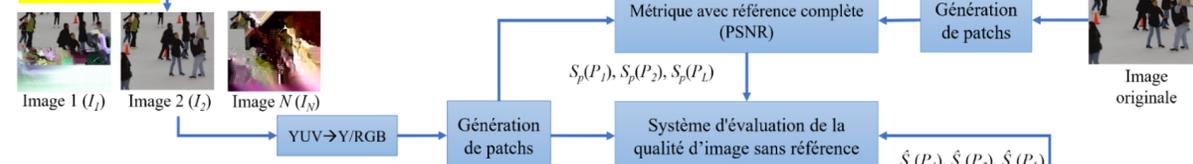
- Méthode basée sur l'usage des architectures d'apprentissage en profondeur afin de permettre la prise en compte de distorsions non-uniformes dues aux erreurs de transmission.
- Une première méthode d'évaluation de la qualité améliorée à plusieurs égards basée sur le CNN présenté dans [2] pour supporter des distorsions non-uniformes dans les images.



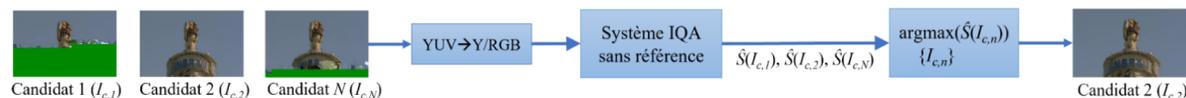
1. Génération de la base de données



2. Entraînement



3. Choisir le meilleur candidat (inférence)



- Une normalisation locale améliorée :

$$v_n(i, j) = \frac{v(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \text{ avec}$$

$$\mu(i, j) = \frac{1}{(2W+1)^2} \sum_{p=-W}^{p=W} \sum_{q=-W}^{q=W} v(i+p, j+q)$$

$$\sigma(i, j) = \sqrt{\frac{1}{(2W+1)^2} \sum_{p=-W}^{p=W} \sum_{q=-W}^{q=W} [v(i+p, j+q) - \mu(i, j)]^2}$$

$$v_n(i, j) = \begin{cases} 0, & \text{si } \sigma(i, j) = 0 \text{ et } \mu(i, j) = 0 \\ \epsilon, & \text{si } \sigma(i, j) = 0 \text{ et } \mu(i, j) \neq 0 \end{cases}$$

- Une mesure de qualité locale opérant par patch :

$$\hat{S}(I) = \frac{1}{L} \sum_{l=0}^{L-1} \hat{S}_p(P_l)$$

Résultats expérimentaux

- Toutes les séquences originales utilisées dans nos expériences proviennent de [3, 4].
- 990 images corrompues à partir de 90 images de référence et finalement 475 200 patches pour l'apprentissage, avec une taille de patch de 64x64 pixels.
- Chaque patch associé à un score PSNR dans l'intervalle [0, 50] dB, qui est normalisé dans l'intervalle [0, 1] pendant l'apprentissage.
- Plusieurs métriques pour évaluer les performances des modèles :

$$\bar{S}_{\text{intact}} = \frac{1}{N} \sum_{n=0}^{N-1} \text{PSNR}_{YUV}(I_{\text{original},n}, I_{\text{intact},n}), \quad \bar{S}_{\text{système}} = \frac{1}{N} \sum_{n=0}^{N-1} \text{PSNR}_{YUV}(I_{\text{original},n}, I_{\text{système},n}), \quad \text{où } I_{\text{système}} = \arg \max_{\{I_{c,i}, 0 \leq i < K\}} \hat{S}(I_{c,i}), \quad \bar{S}_{\text{diff}} = |\bar{S}_{\text{intact}} - \bar{S}_{\text{système}}|$$

Méthodes	Précision	\bar{S}_{intact} (dB)	$\bar{S}_{\text{système}}$ (dB)	\bar{S}_{diff} (dB)
CNN_Y_G pré-entraîné [6]	45.6%	39.18	28.69	10.49
CNN_Y proposé	93.0%		38.39	0.79
CNN_RGB proposé	96.5%		38.88	0.30
CNN_Y_NL proposé	98.2%		38.60	0.58
CNN_RGB_NL proposé	96.5%		38.88	0.30

TABLEAU 1 – Performances des méthodes sur images intra

Méthodes	Précision	\bar{S}_{intact} (dB)	$\bar{S}_{\text{système}}$ (dB)	\bar{S}_{diff} (dB)
CNN_Y_G pré-entraîné [6]	33.3%	38.62	32.99	5.63
CNN_Y proposé	60.0%		30.19	8.43
CNN_RGB proposé	66.7%		36.49	2.13
CNN_Y_NL proposé	77.0%		36.55	2.07
CNN_RGB_NL proposé	79.0%		36.71	1.91

TABLEAU 2 – Performances des méthodes sur images inter

Conclusion / Travaux futurs

- Une méthode utilisée pour extraire la meilleure reconstruction d'une liste d'images candidates générées à l'issue d'un processus de correction d'erreurs.
- Une précision de décision de plus de 98% lorsque l'erreur de transmission est localisée dans l'image intra et une précision de décision d'environ 80% avec l'image inter.
- Ajouter un mécanisme de sélection par patch en référence aux voisins pour améliorer les résultats.

Bibliographie

- [1] Vivien Bousard et al. Table-free multiple bit-error correction using the CRC syndrome. IEEE Access, 8 :102357–102372, 2020.
- [2] Le Kang et al. Convolutional neural networks for no-reference image quality assessment. Dans 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 1733–1740, 2014.
- [3] Xiph.org video test media [derf's collection].
- [4] Margaret H. Pinson. The consumer digital video library [best of the web]. IEEE Signal Processing Magazine, 30(4) :172–174, 2013.
- [5] Yujing ZHANG et al. Optimisation du décodage par liste de vidéos corrompues basée sur une architecture CNN. Accepté par la conférence Compression et Représentation des Signaux Audiovisuels 2023, Lille, 2023.

