Proposition de sujet de thèse 2025 – financement MESR/AAP Région NA

Codage pour machines et humains pour les données visuelles

Résumé et Contexte

Avec l'explosion des données visuelles, les méthodes de compression d'images et de vidéos doivent s'adapter à des besoins variés : maximiser la compression tout en minimisant la perte de qualité perçue par les humains, et garantir des représentations compactes mais exploitables par les machines. Jusqu'à présent, la recherche a exploré deux axes distincts : la compression pour l'Humain, basée sur la perception visuelle, et la compression pour la Machine, optimisée pour les tâches de vision artificielle. Cependant, ces deux paradigmes sont souvent développés séparément et reposent sur des objectifs contradictoires. Quelques travaux ont cherché à développer des approches hybrides combinant les deux aspects. Les avancées récentes en apprentissage profond et en modélisation perceptuelle ouvrent la voie à une compression hybride, capable de s'adapter dynamiquement aux besoins spécifiques des machines et des humains.

État de l'art et limites actuelles

Les méthodes de compression d'images et de vidéos se sont historiquement divisées en deux grandes catégories : celles optimisées pour la perception humaine et celles dédiées aux algorithmes de vision artificielle. Dans le premier cas, les codecs classiques comme JPEG ou H.26x, par exemple, reposent sur des mécanismes visant à maximiser la fidélité visuelle tout en minimisant la quantité de données stockées ou transmises. Ces approches peuvent être optimisées en exploitant des principes issus de la psychologie cognitive, comme le Just Noticeable Difference (JND) ou le Satisfied User Ratio (SUR) [6], afin d'éliminer les informations considérées comme non essentielles par le système visuel humain [5]. Toutefois, ces codecs ne tiennent pas compte des besoins des modèles d'intelligence artificielle (IA), ce qui peut introduire des artefacts indésirables qui perturbent l'analyse automatique des images et vidéos.
En parallèle, la compression pour les machines a récemment émergé comme un champ de recherche crucial, en raison de l'essor des algorithmes de vision par ordinateur et de l'intelligence artificielle appliquée aux images. Des standards comme le Video Coding for Machines (VCM), récemment promu par MPEG [1][3], visent à optimiser directement la compression des vidéos pour des tâches d'analyse automatisée telles que la classification, la segmentation et la détection d'objets. De plus, l'utilisation d'autoencodeurs variationnels (VAE), de modèles génératifs adversaires (GANs) et de transformers visuels [7] a permis de produire des représentations latentes compactes tout en maintenant une performance élevée pour ces tâches [9]. Cependant, ces méthodes souffrent d'un manque d'interprétabilité, car elles génèrent des représentations qui ne sont pas nécessairement lisibles par les humains et qui ne permettent pas une reconstruction fidèle de l'image originale.
Face à ces limites, une nouvelle direction de recherche consiste à unifier ces deux approches en développant des modèles de compression hybrides capables de s'adapter dynamiquement aux besoins des machines et des humains. Des méthodes récentes, comme TransTIC [4] ont tenté de transférer des codecs optimisés pour la perception humaine vers des tâches de vision artificielle sans nécessiter de réentraînement. Par ailleurs, les modèles de compression générative [2] basés sur les modèles de diffusion et les GANs offrent des perspectives intéressantes en permettant une reconstruction optimisée des images selon le type d'utilisateur. Toutefois, ces approches restent computationnellement coûteuses et ne sont pas adaptées aux contraintes énergétiques des systèmes embarqués.
Une autre avancée prometteuse repose sur l'intégration de l'informatique neuromorphique

dans les systèmes de compression. Inspirées des réseaux neuronaux biologiques, ces architectures permettent une gestion dynamique de l'information et une transmission adaptative en fonction des stimuli visuels et des besoins computationnels. En particulier, les réseaux de neurones à impulsions (SNNs) ont démontré un potentiel considérable pour la compression énergétique et la transmission événementielle des données. De plus, des modèles de codage hiérarchique bio-inspirés, s'inspirant du fonctionnement du cortex visuel humain, pourraient permettre une compression adaptative, où l'allocation des ressources serait modulée en fonction de l'importance perceptuelle et des exigences analytiques.

Objectifs de la thèse

L'objectif de cette thèse est de concevoir un modèle de compression hybride, capable de répondre simultanément aux exigences des systèmes de vision artificielle et aux besoins de perception humaine, tout en garantissant une explicabilité et une adaptabilité dynamiques. Contrairement aux approches traditionnelles, qui optimisent séparément la compression pour les machines ou pour les humains, ce projet vise à unifier ces deux paradigmes en exploitant les avancées récentes en traitement du signal, en vision par ordinateur et en intelligence artificielle explicable (XAI).
L'approche proposée s'appuiera sur des modèles d'apprentissage profond et des stratégies de codage adaptatif, intégrant des mécanismes d'explication intégrés pour rendre les représentations compressées interprétables tant par les utilisateurs humains que par les algorithmes de vision artificielle. En combinant des modèles attentionnels explicables et des stratégies de régularisation perceptuelle, cette solution devra non seulement d'améliorer la fidélité visuelle et l'efficacité du codage pour les tâches automatiques, mais aussi de mieux comprendre et contrôler les pertes d'information induites par la compression.
Enfin, une ouverture vers les architectures neuromorphiques permettrait d'explorer des circuits inspirés du cerveau humain pour réduire la consommation énergétique et améliorer la scalabilité computationnelle de la compression. Ces architectures bio-inspirées, en particulier les réseaux de neurones à impulsions (SNNs) et les mémoires associatives adaptatives, offriront de nouvelles perspectives pour un traitement économe en ressources, adapté aux contraintes des systèmes embarqués et des flux vidéo massifs.

Objectifs de la thèse

[1] Zhang, Q., Mei, J., Guan, T., Sun, Z., Zhang, Z., & Yu, L. (2024). Recent Advances in Video Coding for Machines Standard and Technologies. ZTE Communications, 22(1), 62.

[2] *Chen, B., Yin, S., Chen, P., Wang, S., & Ye, Y. (2024). Generative Visual Compression: A Review. arXiv preprint arXiv:2402.02140.*

[3] Lee, D., Jeon, S., Jeong, Y., Kim, J., & Seo, J. (2023). Exploring the Video Coding for Machines Standard: Current Status and Future Directions. Journal of Broadcast Engineering , 28(7), 888-903.

[4] Chen, Y. H., Weng, Y. C., Kao, C. H., Chien, C., Chiu, W. C., & Peng, W. H. (2023). Transtic: Transferring transformer-based image compression from human perception to machine perception. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 23297-23307).

[5] Ballé, J., Laparra, V., & Simoncelli, E. P. (2016). End-to-end optimized image compression. arXiv preprint arXiv:1611.01704.

[6] Zhang, Q., Wang, S., Zhang, X., Jia, C., Wang, Z., Ma, S., & Gao, W. (2024). Perceptual Video Coding for Machines via Satisfied Machine Ratio Modeling. IEEE Trans. on Pattern Analysis and Machine Intelligence.

[7] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[8] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. IEEE Int. Conf. on computer vision.

[9] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. 22nd ACM SIGKDD int. Conf. on knowledge discovery and data mining (pp. 1135-1144).

[10] Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2020). Learned image compression with discretized gaussian mixture likelihoods and attention modules. IEEE/CVF conf. on computer vision and pattern recognition.