

REDES NEURAIS CONVOLUCIONAIS COMO ESTRATÉGIA PARA REDUÇÃO NO TEMPO DE EXECUÇÃO DA DECISÃO DE MODO INTRA NO VVC

ADSON DUARTE¹; BRUNO ZATT²; GUILHERME CORREA³;
DANIEL PALOMINO⁴

¹Universidade Federal de Pelotas – airduarte@inf.ufpel.edu.br

²Universidade Federal de Pelotas – zatt@inf.ufpel.edu.br

³Universidade Federal de Pelotas – gcorrea@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas – dpalomino@inf.ufpel.edu.br

1. INTRODUÇÃO

Aplicações de vídeos digitais têm estado em constante crescimento nos últimos anos, tais como os serviços de vídeo por demanda. Um relatório estatístico mostrou que, ao considerar o mundo inteiro no mês de agosto de 2021, em apenas um minuto de internet existem aproximadamente 694 horas de vídeos transmitidas pelo *YouTube* (DIXON, 2021). Além disto, um estudo feito pela Cisco (2020) estima que 66% dos televisores instalados com acesso a aplicações de vídeo por demanda serão de resolução *Ultra-high-definition* (UHD) até o fim de 2023. Desta forma, é de extrema relevância o estudo de técnicas de codificação de vídeo de ponta, tais como as utilizadas no *Versatile Video Coding Standard* (VVC) (BROSS et al., 2020).

Com o objetivo de aumentar a eficiência de codificação quando comparado a outros padrões de codificação de vídeo, foram desenvolvidas diversas ferramentas de codificação na predição intra do VVC. Entretanto, estas ferramentas também contribuíram para um aumento na complexidade da decisão de modo intra no VVC (SALDANHA et al., 2020). A decisão de modo intra consiste em avaliar as diversas combinações entre tamanhos de bloco e modos intra para uma região do vídeo, onde ao final é escolhida a combinação que resulta no melhor custo taxa-distorção.

Redes Neurais Convolucionais (RNC) têm sido amplamente aplicadas em problemas de classificação de imagens, devido ao fato de que a arquitetura de uma RNC automatiza o processo de extração de características de uma imagem, tornando-a muito robusta ao lidar com este tipo de problema. O problema de otimização presente na decisão de modo intra pode ser modelado como um problema de classificação de imagens, onde a região do vídeo atual na decisão de modo pode servir como entrada para uma RNC, e a saída da RNC consiste em uma classe que indica qual ou quais modo(s) intra são apropriados para a região atual do vídeo.

Existem trabalhos que buscam reduzir a complexidade na decisão de modo intra do VVC baseados em RNCs (HOANGVAN et al., 2021; TANG et al., 2020; TISSIER et al., 2020; ZHANG et al., 2021). Entretanto, estes trabalhos buscam reduzir a quantidade de modos de particionamento avaliados na decisão de modo intra, e não a quantidade de modos intra. Por outro lado, os trabalhos que reduzem a quantidade de modos intra são baseados em algoritmos heurísticos (CHEN et al., 2020; DONG et al., 2022; PARK; KIM; JEON, 2020; ZHANG et al., 2020).

Portanto, o objetivo deste trabalho é utilizar uma RNC como estratégia para reduzir a quantidade de modos avaliados na decisão de modo intra do VVC. Uma análise foi realizada, a qual demonstrou que o modo intra Planar possui uma alta taxa de ocorrência. Desta forma, o problema de classificação foi modelado de forma que a partir da imagem de entrada, a RNC busque classificar se a imagem deve ser predita com o modo intra Planar ou deve ser predita com outros modos intra.

2. METODOLOGIA

A Figura 1 apresenta a taxa de ocorrência dos modos intra no VVC ao agrupar estes em quatro classes: Planar, DC, Angular e *Matrix-based intra prediction* (MIP). Os dados utilizados são de um trabalho prévio (DUARTE, 2021), o qual considera 15 vídeos, 17 tamanhos de bloco, os *Quantization Parameters* (QP) 22, 27, 32 e 37 e o *VVC Test Model* (VTM) 9.0 com a configuração *All Intra*. Na Figura 1, é possível notar que o modo intra Planar possui a maior taxa de ocorrência para todos os QPs. Dada a sua alta taxa de ocorrência, a decisão de modo intra foi modelada como um problema de classificação de imagens binário, onde a RNC recebe como entrada a imagem avaliada pela decisão de modo e, a partir desta, realiza a classificação, onde a classe “planar” indica que o modo planar é adequado para a imagem, e a classe “não planar” indica que o modo planar não é adequado para a imagem.

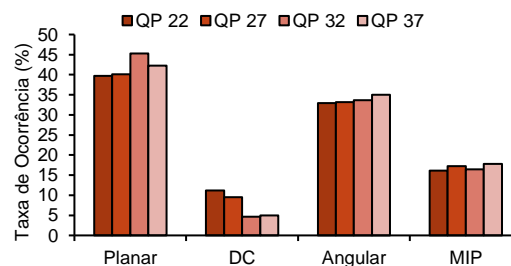


Figura 1: Taxa de ocorrência dos modos intra no VVC por QP.

Um conjunto de dados com diversas imagens e suas respectivas classes foi obtido ao considerar os dados previamente extraídos em Duarte, 2021. Entretanto, nestes dados os exemplos consideram apenas as decisões de modo e características tabulares, isto é, não consideram imagens. Portanto, a partir do vídeo, do número do quadro, das posições x e y e do tamanho de bloco associados a cada exemplo, extraiu-se as imagens a partir dos quadros de cada vídeo. Esta etapa ocorreu no MATLAB e um processo de conversão das imagens do formato YUV para o formato RGB foi realizado. Existem cerca de 200 mil exemplos balanceados por classe, QP, vídeo e tamanho de bloco, e somente os tamanhos de bloco quadrados foram considerados. A classe “planar” contém exemplos onde o modo Planar foi escolhido pela decisão de modo, enquanto a classe “não planar” contém exemplos onde um modo DC, Angular ou MIP foi escolhido pela decisão de modo.

A arquitetura escolhida foi a *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e a biblioteca utilizada foi a *fastai*. Foram selecionados 10% dos dados para compor o conjunto de validação. O treino considerou 50 épocas, *transfer learning* e as funções de perda e otimização *CrossEntropyLoss* e *AdamW*, respectivamente. As características tabulares foram concatenadas na entrada da primeira camada totalmente conectada. Existem um total de 23 características, sendo estas QP, largura, altura e posições x e y do bloco, profundidades da estrutura de particionamentos e das árvores quaternária e binária, menores custos taxa-distorção obtidos no *Rough Mode Decision* para os modos Planar, DC, Angular e MIP, o número *Multiple Reference Line* para os modos DC e Angular, um booleano que indica se o modo MIP é *transposed*, a posição do primeiro modo Planar, DC, Angular e MIP e o número do primeiro modo Angular e MIP na lista de modos a serem avaliados pela decisão de modo, e, por fim, o custo taxa-distorção obtido para o modo Planar no *Rate-Distortion Optimization Process*. As imagens foram redimensionadas para 224x224 e tanto estas quanto as características foram normalizadas.

3. RESULTADOS E DISCUSSÃO

Na Tabela 1 são apresentados os resultados de precisão, revocação e F1 ao considerar as predições da RNC no conjunto de validação, o qual contém 8899 e 8971 exemplos das classes “não planar” e “planar”, respectivamente. As métricas de precisão, revocação e F1 variam no intervalo $[0,1]$, onde valores próximos de um são melhores. A precisão indica, para cada classe, a razão entre a quantidade de exemplos que foram classificados corretamente e a quantidade total de exemplos classificados na classe. A revocação complementa a precisão e indica, para cada classe, a proporção de exemplos que pertencem a classe e que foram classificados como tal. Por fim, a F1 é uma média harmônica entre precisão e revocação.

Ao observar os resultados na Tabela 1, nota-se que a RNC possui uma precisão um pouco maior para a classe “não planar”, o que em contrapartida resulta em uma revocação maior para a classe “planar”. Isto indica que a RNC tem uma tendência, ainda que baixa, de classificar exemplos da classe “não planar” como sendo da classe “planar”. Por outro lado, quando um exemplo é classificado na classe “não planar” pela RNC, a precisão é maior, isto é, as chances deste exemplo pertencer de fato a classe “não planar” são mais altas quando comparadas a classe “planar”. Ainda assim, por se tratar de diferenças pequenas, a F1 para ambas as classes foi muito semelhante, com uma média de 0,72.

Tabela 1: Resultados de precisão, revocação e F1 por classe para a RNC.

Classe	Precisão	Revocação	F1	Suporte
não planar	0,73	0,70	0,71	8899
planar	0,71	0,74	0,72	8971
Média	0,72	0,72	0,72	

4. CONCLUSÕES

Uma RNC foi proposta como estratégia para reduzir a quantidade de modos intra avaliados na decisão de modo do VVC. Dada a alta taxa de ocorrência do modo Planar, o problema de decisão de modo foi modelado como um problema de classificação de imagens binário, onde o objetivo da RNC é prever, para cada região do vídeo, as classes que indicam se a região deve ser predita com o modo planar ou não deve ser predita com o modo planar. Os próximos passos consistem em integrar a RNC obtida no software de referência do VVC para obter os resultados de tempo de execução e eficiência de codificação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BROSS, B.; CHEN, J.; LIU, S.; WANG, Y.-K. Versatile Video Coding Editorial Refinements on Draft 10. **JVET-T2001**, v. 2. 2020. Disponível em: https://jvet-experts.org/doc_end_user/current_document.php?id=10540. Acesso em: 15, ago. 2022.

CHEN, Y. et al. A novel fast intra mode decision for versatile video coding. **Journal of Visual Communication and Image Representation**, China, v.71, p.102849, 2020.

CISCO. Cisco Annual Internet Report (2018-2023) White Paper. **CISCO**, 2020. Disponível em: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. Acesso em: 10, ago. 2022.

DUARTE, A. **Redução de complexidade do processo de decisão de modo da predição intra-quadro do codificador de vídeo VVC utilizando aprendizado de máquina**. 2021. 107f. Dissertação (Mestrado em Computação) – Curso de Pós-Graduação em Computação, Universidade Federal de Pelotas.

DIXON, S. Media usage in an internet minute as of August 2021. **Statista**, 2021. Disponível em: <https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/>. Acesso em: 10, ago. 2022.

DONG, X.; SHEN, L.; YU, M.; YANG, H. Fast Intra Mode Decision Algorithm for Versatile Video Coding. **IEEE Transactions on Multimedia**, China, v.24, p.400–414, 2022.

HOANGVAN, X. et al. Fast QTMT for H. 266/VVC Intra Prediction using Early-Terminated Hierarchical CNN model. In: INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES FOR COMMUNICATIONS (ATC), 2021., 2021. **Anais. . . IEEE**, 2021. p.195–200.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, Canada, v.25, 2012.

PARK, J.; KIM, B.; JEON, B. Fast VVC intra prediction mode decision based on block shapes. In: APPLICATIONS OF DIGITAL IMAGE PROCESSING XLIII, 2020. **Anais. . . SPIE**, 2020. v.11510, p.581 – 593.

SALDANHA, M.; SANCHEZ, G.; MARCON, C.; AGOSTINI, L. Complexity Analysis Of VVC Intra Coding. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP), 2020., 2020. **Anais. . . IEEE**, 2020. p.3119–3123.

TANG, G.; JING, M.; ZENG, X.; FAN, Y. Adaptive CU split decision with poling-variable CNN for VVC intra encoding. In: IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), 2019., 2019. **Anais. . . IEEE**, 2019. p.1–4.

TISSIER, A. et al. CNN oriented complexity reduction of VVC intra encoder. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP), 2020., 2020. **Anais. . . IEEE**, 2020. p.3139–3143.

ZHANG, Q.; GUO, R.; JIANG, B.; SU, R. Fast CU decision-making algorithm based on DenseNet network for VVC. **IEEE Access**, China, v.9, p.119289–119297, 2021.

ZHANG, Q.; WANG, Y.; HUANG, L.; JIANG, B. Fast CU partition and intra mode decision method for H. 266/VVC. **IEEE Access**, China, v.8, p.117539–117550, 2020.