

## ACELERAÇÃO DA PREDIÇÃO INTRA-QUADRO DO PADRÃO DE CODIFICAÇÃO DE VÍDEO VVC UTILIZANDO REDES NEURAIS

ENTHONY BOHM<sup>1</sup>; ADSON DUARTE<sup>2</sup>; DANIEL PALOMINO<sup>3</sup>

<sup>1</sup>Universidade Federal de Pelotas – [egrbohm@inf.ufpel.edu.br](mailto:egrbohm@inf.ufpel.edu.br)

<sup>2</sup>Universidade Federal de Pelotas – [airduarte@inf.ufpel.edu.br](mailto:airduarte@inf.ufpel.edu.br)

<sup>3</sup>Universidade Federal de Pelotas – [dpalomino@inf.ufpel.edu.br](mailto:dpalomino@inf.ufpel.edu.br)

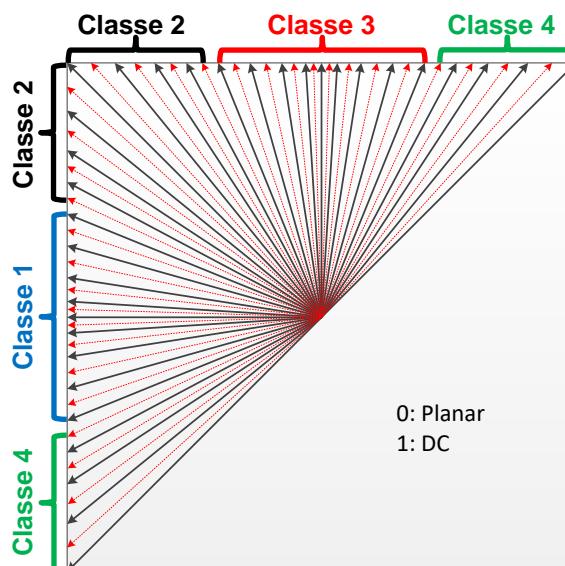
### 1. INTRODUÇÃO

O consumo de vídeos vem crescendo ao longo dos anos. Segundo o relatório da Internet Sandvine (2023), na primeira metade de 2022, aplicações de vídeos representaram 65% do tráfego online. Logo, o estudo de padrões de codificação de vídeo eficientes, como o *Versatile Video Coding* (VVC) (BROSS et al, 2020), torna-se de extrema relevância.

O VVC é um padrão estado da arte que atinge uma compressão 40% superior ao *High Efficiency Video Coding* (HEVC), embora exija um esforço computacional 31 vezes maior (PAKDAMAN, 2020). Esse aumento está relacionado às novas ferramentas de codificação intra, como a estrutura de particionamento e a ampliação de 33 para 66 modos angulares (BROSS et al, 2020). Esses novos modos melhoram a eficiência, mas também aumentam o esforço computacional, devido ao aumento no número de combinações avaliadas pelo processo *Rate Distortion Optimization* (RDO) (SULLIVAN, 1998). Embora o software de referência VVC Test Model (VTM) avalie apenas um subconjunto de modos (*RD-List*), o número de modos avaliados pelo RDO permanece elevado.

Desta forma, o objetivo deste trabalho é propor uma solução baseada em Redes Neurais Convolucionais (RNC) para reduzir a quantidade de modos angulares avaliados no RDO. Os modos angulares foram agrupados em quatro classes de acordo com as suas direções, conforme a Figura 1, e uma RNC baseada na LeNet-5 (LECUN, 1998) foi treinada para predizer a classe mais apropriada para cada bloco. Com isso, os modos fora da classe predita são removidos da *RD-List*, diminuindo o número de iterações necessárias no RDO.

Figura 1- Modos angulares do VVC agrupados em quatro classes



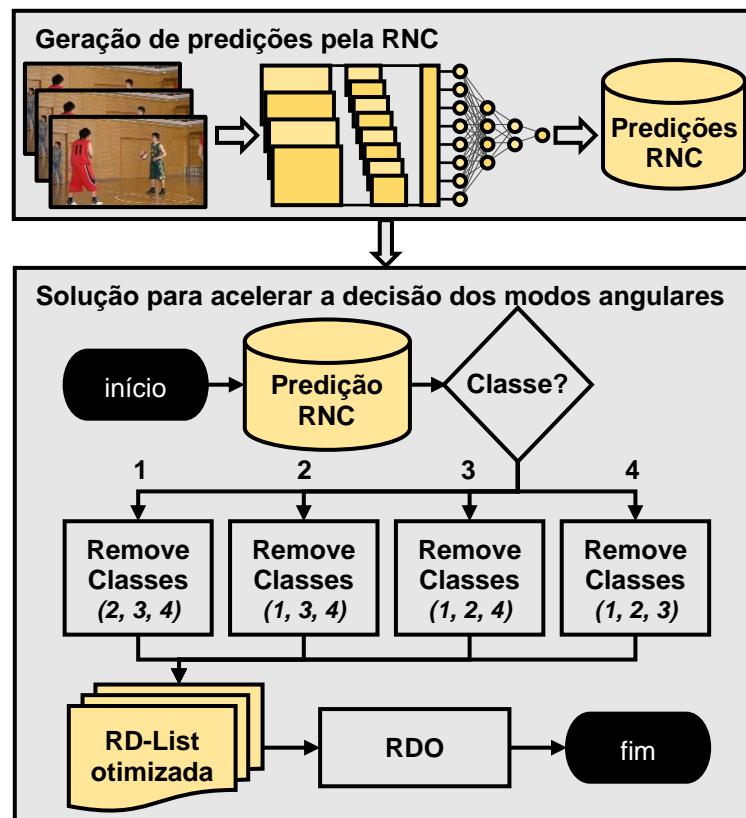
## 2. METODOLOGIA

A Figura 2 apresenta a solução proposta neste trabalho. A RNC gera previsões para cada bloco, as quais são acessadas pela decisão de modo intra do VTM. De acordo com a classe predita pela RNC, a solução remove da *RD-List* os modos angulares que não pertencem a classe predita pela RNC, resultando em uma *RD-List* otimizada. Como a *RD-List* otimizada possui uma quantidade menor de modos angulares, consequentemente o número de iterações necessárias no RDO para encontrar o melhor modo é menor, o que reduz o esforço computacional.

O conjunto de dados utilizado para o treinamento da RNC foi obtido ao seguir a metodologia de Duarte et al. (2023), onde 15 vídeos com resoluções 4K, 1080p e 720p foram codificados no VTM 21.2 com a configuração *All Intra* e os valores de QP (*Quantization Parameter*) 22, 27, 32 e 37. Destas codificações extraiu-se, para cada bloco processado, a decisão de modo intra final, e os exemplos foram então agrupados nas quatro classes apresentadas na Figura 1. O conjunto de dados contém cerca de 3 milhões de exemplos balanceados por vídeo, tamanho de bloco, QP e classe.

A arquitetura Lenet-5 foi utilizada, e o treinamento foi conduzido por 100 épocas com a biblioteca *Pytorch* em uma GPU RTX3080. A função de perda e o otimizador utilizados foram *cross entropy* e *Adam*, respectivamente, com um *learning rate* de 0,001. Para o treinamento, 90% dos exemplos foram utilizados, enquanto os 10% dos exemplos restantes foram reservados para validar cada época. Ao final das 100 épocas, selecionou-se o modelo que obteve a menor perda no conjunto de validação, o qual resultou em 79% de acurácia.

Figura 2 – Solução para acelerar a decisão de modo intra dos modos angulares no VVC



### 3. RESULTADOS E DISCUSSÃO

Para avaliar o desempenho da solução proposta, a RNC obtida foi utilizada para gerar previsões para três vídeos das *Common Test Conditions* (BOSSEN et al., 2020) do VVC, sendo estes *Foodmarket4* da Classe A1 (4K), *BasketballDrive* da classe B (1080p) e *FourPeople* da Classe E (720p). É importante ressaltar que **nenhum** destes vídeos foi utilizado para treinar a RNC. Após, cada vídeo foi codificado com os quatro valores de QP 22, 27, 32 e 37 no VTM 21.2 original e no VTM 21.2 com a solução proposta. A partir destas codificações, duas métricas são medidas: redução de tempo (RT), a qual indica o percentual de redução no tempo de codificação obtido pela solução proposta, e BDBR (BJONTEGAARD, 2001), a qual mede a perda em eficiência de codificação inserida pela solução.

A Tabela 1 apresenta os resultados de redução de tempo (RT) e de eficiência de codificação (BDBR). Como pode ser visto, a solução proposta obtém em média 4,07% de redução no tempo de codificação com apenas 0,21% de perda em eficiência de codificação. Para o vídeo *FourPeople*, a solução obtém o melhor resultado de redução no tempo de codificação, atingindo 4,44% com apenas 0,21% de perda em eficiência de codificação.

Tabela 1 – Resultados de redução de tempo de codificação e BDBR

Classe	Vídeo	RT	BDBR
A1	FoodMarket4	4,02%	0,12%
B	BasketballDrive	3,76%	0,23%
E	FourPeople	4,44%	0,30%
<b>Média</b>		<b>4,07%</b>	<b>0,21%</b>

A Tabela 2 apresenta uma comparação com trabalhos relacionados. Para comparação das soluções foram consideradas as médias de RT, BDBR e a razão RT/BD-BR, que indica o ganho em tempo de codificação para cada 1% que se perde na eficiência de compressão. Através dos resultados, constata-se que apesar da solução proposta obter uma redução no tempo de codificação menor, por focar apenas nos modos angulares, ela também oferece a menor perda em eficiência de codificação, representada pelo BD-BR, apresentando resultados competitivos no *tradeoff* entre redução no tempo e perda em eficiência de codificação (RT/BDBR).

Tabela 2 – Comparação com trabalhos relacionados

Solução	Software	RT	BD-BR	RT/BDBR
<b>Solução Proposta</b>	<b>VTM 21.2</b>	<b>4,07%</b>	<b>0,21%</b>	<b>21,77</b>
Liu et al. (2023)	VTM 14.0	16,72%	0,83%	20,14
Saldanha et al. (2021)	VTM 10.0	10,47%	0,29%	36,10

### CONCLUSÕES

Neste trabalho, foi apresentada uma solução baseada em redes neurais convolucionais para acelerar a decisão de modo intra dos modos angulares no VVC. Os modos angulares são agrupados em quatro classes e uma Lenet-5 prediz a classe mais provável para cada bloco, evitando a avaliação dos modos angulares pertencentes às demais classes. A solução obtém bons resultados de redução no tempo de codificação com baixas perdas em eficiência de codificação, atingindo resultados competitivos em comparação a trabalhos relacionados.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- BJONTEGAARD, G. **Calculation of average PSNR differences between RD-Curves.** VCEG Meeting, [https://www.itu.int/wftp3/av-arch/video-site/0104\\_Aus/VCEG-M33.doc](https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc).
- BOSSEN, F. et al. **VTM common test conditions and software reference configurations for SDR video.** JVET-T2010-v1, [https://jvetexperts.org/doc\\_end\\_user/current\\_document.php?id=10545](https://jvetexperts.org/doc_end_user/current_document.php?id=10545).
- BROSS B. et al., Overview of the Versatile Video Coding (VVC) Standard and its Applications, **IEEE Transactions on Circuits and Systems for Video Technology**, vol. 31, no. 10, 3736-3764, 2021
- DUARTE, A. et al., Fast Intra Mode Decision Using Machine Learning for the Versatile Video Coding Standard, **IEEE International Symposium on Circuits and Systems (ISCAS)**, Monterey, CA, USA, 1–5, 2023.
- LECUN Y. et al., Gradient-based learning applied to document recognition, **Proceedings of the IEEE**, vol. 86, no. 11, 2278-2324, Nov. 1998.
- LIU, Z. et al., Deep Multi-task Learning based Fast Intra-mode Decision for Versatile Video Coding, **IEEE Transactions on Circuits and Systems for Video Technology**, 1–1, 2023.
- PAKDAMAN, F. et al., Complexity Analysis of Next-Generation VVC Encoding and Decoding, **IEEE International Conference on Image Processing (ICIP)**, Abu Dhabi, United Arab Emirates, 3134-3138, 2020
- SALDANHA, M. et al., Learning-Based Complexity Reduction Scheme for VVC Intra-Frame Prediction, **International Conference on Visual Communications and Image Processing (VCIP)**, Munich, Germany, 1-5, 2021.
- SANDVINE, **Sandvine's 2023 Global Internet Phenomenal Report**, Janeiro 2023. Disponível em: [https://www.sandvine.com/hubfs/Sandvine\\_Redesign\\_2019/Downloads/2023/reports/Sandvine%20GIPR%202023.pdf](https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2023/reports/Sandvine%20GIPR%202023.pdf)
- SULLIVAN G. J.; WIEGAND T., Rate-distortion optimization for video compression, **IEEE Signal Processing Magazine**, vol. 15, no. 6, 74-90, Nov.1998