APJ. 6(1):55-66, Spring 2025

Arman Process Journal (APJ)

Quarterly Journal of ICT APJ License Number: 87090

Arman Process Journal (APJ)

Homepage: https://www.armanprocessjournal.ir

Optimizing Video Coding Using Neural Networks: A Comprehensive Review of Methods and Applications

***Mehran Riki* [*,1], *Fatemeh Mohammadi* [2] , *Pouria Khazeni*[3]**

[1] *Faculty Member, Department of Computer Engineering, National University of Skill(NUS), Tehran, Iran*

[2] *Master's Student in Computer Science, Faculty of Mathematics, Statistics, and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran*

[3] *Master's Student in Information Technology Management, Faculty of Management and Economics, University of Sistan and Baluchestan, Zahedan, Iran*

RESEARCH PAPER

Received: 11 December 2024
Accepted: 21 April 2025

KEYWORDS:
Video Coding,
Neural Networks,
Optimization,
Video Hiding,
Deep Learning.

[1] Corresponding author:
mriki@nus.ac.ir

ABSTRACT

Background and Objectives: With the rising demand for high-quality video services such as live streaming, virtual reality, and UHD videos, efficient video coding methods have become increasingly critical. Standards like H.264, H.265, and H.266 aim to reduce data volume while preserving image quality, yet they face challenges such as computational complexity, bandwidth management, and compression resilience. This paper provides a comprehensive review of neural network-based approaches for optimizing video coding. The reviewed methods include convolutional neural networks (CNNs) for intra-mode prediction, LSTM and Seq2Seq models for video traffic modeling, and inverse neural networks (INNs) for robust video hiding. Findings indicate that these techniques can reduce encoding time by up to 70%, enhance frame size prediction accuracy to 13.6%, and improve the resilience of stego videos against compression. Practical applications in bandwidth management, coding optimization for resource-constrained devices, and video security are explored. This study underscores the significant potential of deep learning in advancing video coding standards and suggests future research directions.

Methods: In this section, neural network-based methods with various approaches are examined and analyzed in detail. First, relevant standards such as H.264, H.265, H.266/FVC, and HEVC are introduced, and then each method is analyzed in terms of performance, advantages, and limitations. For example, in the method using CNN for intra prediction, video frames are divided into 16×16 blocks, and two convolutional layers are utilized along with activation functions like ReLU to select only 5 optimal states from 67 states. The neural network-based methods in this study are categorized based on video coding standards and their objectives: 1. Intra prediction with CNN (H.266/FVC): - Method: Using CNN with two convolutional layers to select 5 optimal states from 67 states. - Advantages: 93% reduction in encoding time with improved BDBR (0.033%). - Disadvantages: Limited to 16×16 blocks and poorer performance at 4K. 2. Traffic modeling with LSTM and Seq2Seq (H.264/H.265): - Method: LSTM for long-term dependencies and Seq2Seq for multi-step prediction. - Advantages: MAPE of 13.6% and a 7.6% reduction in inconsistent traffic. - Disadvantages: Requires diverse data for dynamic videos. 3. Video hiding with VHNet (H.264/H.265): - Method: INN and CompressionSimulator for resilience against compression. - Advantages:

PSNR of 33.7 decibels and high generalizability. - Disadvantages: Fixed division into 8 frames regardless of frame types.

Findings: In the field of video coding optimization, numerous practical applications of neural models can be observed in real-world scenarios. For instance, in bandwidth management in wireless networks, Seq2Seq models, by predicting frame sizes, can enhance dynamic bandwidth allocation; a 7.6 percent reduction in inconsistent traffic in VBR networks indicates their successful performance.

Conclusion: This article examines the comprehensive optimization of video coding using neural networks. The findings indicate that models based on CNN, LSTM, Seq2Seq, VHNet, and IFFT-LSTM have played a crucial role in improving the performance of coding systems by reducing encoding time by over 70%, enhancing the accuracy of frame size prediction by 13.6%, and increasing resilience against compression. The diverse applications of these methods in areas such as bandwidth management in wireless networks, encoding optimization in resource-constrained devices, and video transmission security demonstrate their high flexibility and applicability in real-world environments. The use of tools like TensorFlow and PyTorch, along with standard datasets such as UCF101 and REDS, has facilitated the practical implementation of these techniques. Despite significant achievements, challenges such as the need for diverse data to train recurrent models and limitations in processing very high resolutions still exist, presenting avenues for future research. Future studies could enhance the performance of coding systems in 8K applications and sixth-generation 6G networks by integrating novel architectures like transformers and utilizing optimized hardware such as FPGAs. Overall, recent advancements in deep learning offer a promising future for improving video coding standards and enhancing the quality of services provided to users.

نشریه تخصصی آرمان پردازش، دوره ۶، شماره ۱، بهار ۱۴۰۴

فصلنامه تخصصی آرمان پردازش

(APJ)

Homepage: www.armanprocessjournal.ir

فصلنامه تخصصی فناوری اطلاعات و ارتباطات
شماره مجوز: ۸۷۰۹۰

بهینه‌سازی کدگذاری ویدئویی با استفاده از شبکه‌های عصبی: مروری جامع بر روش‌ها و کاربردها

مهران ریکی*[۱]، فاطمه محمدی[۲] ، پوریا خازنی[۳]

[۱] هیات علمی گروه مهندسی کامپیوتر،دانشگاه ملی مهارت،تهران،ایران

[۲] دانشجو کارشناسی ارشد علوم کامپیوتر ، دانشکده ریاضی،آمار و کامپیوتر ، دانشگاه سیستان و بلوچستان،زاهدان،ایران

[۳] دانشجو کارشناسی ارشد مدیریت فناوری اطلاعات،دانشکده مدیریت و اقتصاد،دانشگاه سیستان و بلوچستان،زاهدان،ایران

چکیده

مقاله پژوهشی

پیشینه و اهداف: با افزایش تقاضا برای خدمات ویدئویی با کیفیت بالا مانند پخش زنده، واقعیت مجازی و ویدئوهای UHD، نیاز به روش‌های کارآمد کدگذاری ویدئویی بیش از پیش احساس می‌شود. استانداردهای کدگذاری مانند H.264، H.265 و H.266 با هدف کاهش حجم داده‌ها و حفظ کیفیت تصویر توسعه یافته‌اند، اما همچنان با چالش‌هایی نظیر پیچیدگی محاسباتی، مدیریت پهنای باند و مقاومت در برابر فشرده‌سازی مواجه هستند. این مقاله مروری جامع بر روش‌های مبتنی بر شبکه‌های عصبی برای بهینه‌سازی کدگذاری ویدئویی ارائه می‌دهد. روش‌های مورد بررسی شامل استفاده از شبکه‌های عصبی کانولوشنی (CNN برای پیش‌بینی حالت‌های درونی، شبکه‌های LSTM و Seq2Seq برای مدلسازی ترافیک ویدئویی و شبکه‌های عصبی معکوس (INN برای پنهان‌سازی مقاوم ویدئو هستند. نتایج نشان می‌دهند که این روش‌ها می‌توانند زمان کدگذاری را تا ۷۰ درصد کاهش دهند، دقت پیش‌بینی اندازه فریم‌ها را تا ۱۳٫۶ در صد بهبود بخ شند و مقاومت ویدئوهای ا ستگو را در برابر ف شرده سازی افزایش دهند. کاربردهای عملی این روش‌ها در مدیریت پهنای باند، بهینه سازی کدگذاری در د ستگاه‌های کم‌منبع و امنیت ویدئویی نیز برر سی شده ا ست. این مطالعه ن شان‌دهنده پتان سیل بالای یادگیری عمیق در ارتقای ا ستانداردهای کدگذاری ویدئویی ا ست و جهت‌گیری‌های آینده را پیشنهاد می‌دهد.

روش‌ها: در این بخش، روش‌های مبتنی بر شبکه‌های عصبی با رویکردهای مختلف مورد بررسی قرار گرفته و به تفصیل تحلیل می‌شوند. ابتدا استانداردهای مربوطه مانند H.264، H.265، H.266/FVC و HEVC معرفی و سپس هر روش از جنبه‌های عملکردی، مزایا و محدودیت‌ها تحلیل می‌شود. به عنوان نمونه، در روش استفاده از CNN برای پیش‌بینی حالت‌های درونی، فریم‌های ویدئویی به بلوک‌های ۱۶×۱۶ تقسیم شده و از دو لایه کانولوشنی به همراه توابع

واژگان کلیدی:
کدگذاری ویدئویی،
شبکه‌های عصبی،
بهینه‌سازی،
پنهان‌سازی ویدئو،
یادگیری عمیق .

فعال سازی مانند ReLU بهره گرفته می شود تا تنها ۵ حالت بهینه از میان ۶۷ حالت انتخاب شود.روش‌های مبتنی بر شبکه‌های عصبی در این مطالعه بر اساس استانداردهای کدگذاری ویدئویی و اهدافشان دسته‌بندی شده‌اند:

۱. **پیش‌بینی حالت درونی با CNN (H.266/FVC):**
   - **روش:** استفاده از CNN با دو لایه کانولوشنی برای انتخاب ۵ حالت بهینه از ۶۷ حالت.
   - **مزایا:** کاهش ۹۳ درصدی زمان کدگذاری با BDBR بهبودیافته (۰,۰۳۳ درصد).
   - **معایب:** محدود به بلوک‌های ۱۶×۱۶ و عملکرد ضعیف‌تر در K.۴
۲. **مدلسازی ترافیک با LSTM و Seq2Seq (H.264/H.265):**
   - **روش:** LSTM برای وابستگی‌های بلندمدت و Seq2Seq برای پیش‌بینی چندمرحله‌ای.
   - **مزایا:** MAPE 13.6 درصد و کاهش ۷,۶ درصدی ترافیک ناسازگار.
   - **معایب:** نیاز به داده‌های متنوع برای ویدئوهای پویا.
۳. **پنهان‌سازی ویدئویی با VHNet (H.264/H.265):**
   - **روش:** INN و CompressionSimulator برای مقاومت در برابر فشرده‌سازی.
   - **مزایا:** PSNR 33.7 دسی‌بل و تعمیم‌پذیری بالا.
   - **معایب:** تقسیم ثابت به ۸ فریم بدون توجه به نوع فریم‌ها.

یافته‌ها: در حوزه بهینه سازی کدگذاری ویدئویی، کاربردهای عملی متعددی از مدل‌های عصبی در سناریوهای واقعی مشاهده می‌شود. به عنوان نمونه در مدیریت پهنای باند در شبکه‌های بی‌سیم، مدل‌های Seq2Seq با پیش‌بینی اندازه فریم‌ها قادرند تخصیص پویای پهنای باند را بهبود بخشند؛ به‌طوری که کاهش ۷,۶ درصدی ترافیک ناسازگار در شبکه‌های VBR نشان از عملکرد موفق آن دارد.

نتیجه‌گیری: این مقاله به بررسی جامع و بهینه‌سازی کدگذاری ویدئویی با استفاده از شبکه‌های عصبی پرداخته شد. یافته‌ها نشان می‌دهد که مدل‌های مبتنی بر CNN، LSTM، Seq2Seq، VHNet و IFFT-LSTM توانسته‌اند با کاهش زمان کدگذاری تا بیش از ۷۰ درصد، بهبود دقت پیش‌بینی اندازه فریم‌ها به میزان ۱۳,۶ درصد و افزایش مقاومت در برابر فشرده‌سازی، نقشی اساسی در بهبود عملکرد سیستم‌های کدگذاری ایفا کنند. کاربردهای متنوع این روش‌ها در حوزه‌هایی مانند مدیریت پهنای باند در شبکه‌های بی‌سیم، بهینه‌سازی کدگذاری در دستگاه‌های کم‌منبع و امنیت انتقال ویدئو، نشان از انعطاف‌پذیری و کاربردپذیری بالای آن‌ها در محیط‌های واقعی دارد. استفاده از ابزارهایی مانند TensorFlow، PyTorch و بهره‌گیری از دیتاست‌های استاندارد نظیر UCF101 و REDS، زمینه پیاده‌سازی عملی این تکنیک‌ها را فراهم ساخته است. با وجود دستاوردهای چشمگیر، چالش‌هایی همچون نیاز به داده‌های متنوع برای آموزش مدل‌های بازگشتی و محدودیت‌هایی در پردازش رزولوشن‌های بسیار بالا همچنان وجود دارد که راهگشای پژوهش‌های آینده محسوب می‌شوند. پژوهش‌های آتی می‌توانند با ادغام معماری‌های نوین مانند ترانسفورمرها و استفاده از سخت‌افزارهای بهینه شده مانند FPGA، عملکرد سیستم‌های کدگذاری را در کاربردهای K۸ و شبکه‌های نسل ششم G۶ بهبود بخشند. به طور کلی، پیشرفت‌های اخیر در یادگیری عمیق، امیدبخش آینده‌ای روشن برای ارتقای استانداردهای کدگذاری ویدئویی و ارتقای کیفیت خدمات ارائه شده به کاربران می‌باشد.

مقدمه

در دهه اخیر، پیشرفت‌های شگفت‌انگیز در حوزه فناوری‌های ویدئویی، از جمله پخش زنده، واقعیت مجازی(VR) ، واقعیت افزوده(AR) ، و ویدئوهای با وضوح فوق‌العاده بالا Ultra-High Definition یا UHD ، تقاضا برای روش‌های پیشرفته‌تر و کارآمدتر کدگذاری ویدئویی را به شدت افزایش داده است. این نیاز نه تنها به دلیل افزایش حجم داده‌های ویدئویی، بلکه به خاطر ضرورت حفظ کیفیت بصری در کنار کاهش پهنای باند مورد نیاز و مصرف انرژی، بیش از پیش احساس می‌شود. استانداردهای کدگذاری ویدئویی مدرن مانند H.264/AVC، H.265/HEVC، و جدیدترین آن‌ها یعنی H.266/VVC، با هدف فشرده سازی بهینه داده‌ها و ارائه کیفیت بالا طراحی شده‌اند، اما با چالش‌هایی جدی مواجه هستند. این چالش‌ها شامل پیچیدگی محاسباتی بالا، نیاز به مدیریت پویای پهنای باند در شبکه‌های متغیر، و مقاومت در برابر افت کیفیت ناشی از فشرده‌سازی است. به عنوان مثال، در استاندارد H.266 ، تعداد حالت‌های درونی (Intra Modes)به ۶۷ افزایش یافته که این امر فرآیند بهینه‌سازی نرخ-اعوجاج Rate-Distortion Optimization یا RDO را به شدت زمان‌بر و پیچیده می‌کند. از سوی دیگر، ترافیک ویدئویی متغیر Variable Bit Rate یا VBR در شبکه‌های بی‌سیم، پیش‌بینی دقیق اندازه فریم‌ها را دشوار کرده و تخصیص بهینه پهنای باند را به چالشی بزرگ تبدیل می‌کند. این مسائل به‌ویژه در دستگاه‌های با منابع محدود مانند گوشی‌های هوشمند، تبلت‌ها، و سیستم‌های اینترنت اشیا (IOT) که نیازمند پردازش سریع، مصرف انرژی کم، و عملکرد پایدار هستند، بیش‌تر برجسته می شوند. در این راستا، شبکه‌های عصبی و یادگیری عمیق به عنوان راه‌حل‌هایی نوآورانه و قدرتمند برای غلبه بر این موانع مطرح شده‌اند و پتانسیل بالایی در بهبود فرآیندهای کدگذاری ویدئویی از خود نشان داده‌اند. برای درک بهتر اهمیت این موضوع، جدول زیر سهم ترافیک ویدئویی در شبکه‌های جهانی و رشد پیش‌بینی‌شده آن را نشان می‌دهد:

جدول1 – سهم ترافیک ویدئویی در شبکه‌های جهانی

رشد سالانه(%)	سهم ترافیک ویدئویی(%)	سال
30	75	2018
25	80	2020
20	82	2023
15 (پیش‌بینی)	85 (پیش‌بینی)	2025

این داده‌ها نشان‌دهنده روند رو به رشد ترافیک ویدئویی و ضرورت توسعه روش های کارآمدتر برای مدیریت و کدگذاری آن است. شبکه‌های عصبی مانند CNN شبکه‌های کانولوشنی ، LSTM حافظه بلند-کوتاه مدت ، Seq2Seq، و INN شبکه‌های عصبی معکوس می‌توانند با یادگیری الگوهای پیچیده و پردازش داده های حجیم، راه‌حل‌هایی برای این چالش‌ها ارائه دهند.

پیشینه و مروری بر ادبیات تحقیق

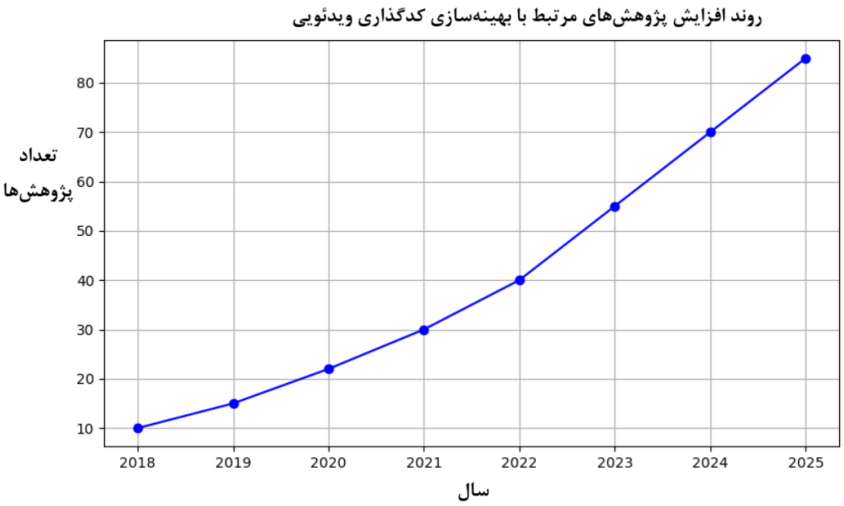
در سال‌های اخیر، پژوهش‌های فراوانی در زمینه بهینه سازی کدگذاری ویدئویی با بهره‌گیری از شبکه‌های عصبی انجام شده است که با هدف کاهش پیچیدگی‌های محاسباتی و بهبود کیفیت فشرده سازی صورت گرفته‌اند. مطالعات متعددی مانند [۱]، [۲] و [۳] از الگوریتم های CNN، LSTM/Seq2Seq و VHNet بهره برده‌اند تا چالش‌های مربوط به پیش‌بینی حالت‌های درونی، ترافیک متغیر و پنهان‌سازی مقاوم ویدئو را حل کنند. به عنوان مثال، استفاده از شبکه های کانولوشنی در استاندارد H.266 منجر به کاهش زمان کدگذاری تا ۹۳ درصد شده و دقت پیش‌بینی با استفاده از مدل‌های بازگشتی مانند LSTM، خطای MAPE را به ۱۳٫۶ درصد کاهش داده است. شکل ۱ نموداری از روند افزایش تعداد مقالات و پروژه‌های تحقیقاتی در این حوزه را نشان می‌دهد. همچنین، جدول ۱ مقایسه‌ای از روش‌های به‌کار رفته از نظر مزایا و محدودیت‌ها ارائه می‌دهد. الگوریتم‌های مطرح در این حوزه معمولاً شامل مراحل پیش‌پردازش داده، آموزش شبکه‌های عمیق، و ارزیابی با معیارهایی مانند PSNR و BD-BR می‌باشند. استفاده از چارچوب‌های نرم‌افزاری مانند TensorFlow و PyTorch نقش مهمی در شتاب‌دهی به این پژوهش‌ها داشته است. علاوه بر این، ترکیب مدل های مختلف در برخی تحقیقات منجر به بهبود عملکرد کلی سیستم‌های کدگذاری شده شده و به کاهش اثرات محدودیت‌های مدل‌های تک‌گانه کمک نموده است. این ادبیات تحقیق

نشـــان می‌دهد که توســـعه روش‌های ترکیبی و اســـتفاده از داده‌های حجیم، زمینه ساز پی شرفت‌های چ شمگیر در ا ستانداردهای کدگذاری ویدئویی آینده خواهد بود.

جدول2 – مقایسه استاندارد ها

روش	استاندارد	مزیت اصلی	محدودیت
CNN	H.266	کاهش ۹۳٪ زمان کدگذاری	محدود به بلوک ۱۶×۱۶
LSTM/Seq2Seq	H.264/H.265	دقت MAPE 13.6%	نیاز به داده‌های متنوع
VHNet	H.264/H.265	PSNR 33.7دسی‌بل	تقسیم‌بندی ثابت فریم‌ها
IFFT-LSTM	HEVC	کاهش ۷۰٪ زمان کدگذاری	افت کیفیت جزئی

این جدول نشان‌دهنده تنوع رویکردها و تعادل بین مزایا و محدودیت‌ها است.



شکل ۱ :نمودار روند افزایش پژوهش‌های مرتبط با بهینه‌سازی کدگذاری ویدئویی از سال ۲۰۱۸ تا کنون.

روش‌شناسی، استانداردها و مقایسه آن‌ها

در این بخش، روش‌های مبتنی بر شـــبکه‌های عصـــبی با رویکردهای مختلف مورد بررســـی قرار گرفته و به تفصـــیل تحلیل می‌شـــوند. ابتدا اســـتانداردهای مربوطه مانند H.264، H.265، H.266/FVC و HEVC معرفی و ســـپس هر روش از جنبه های عملکردی، مزا یا و محدودیت‌ها تحلیل می‌شـــود. به عنوان نمونه، در روش اســـتفاده از CNN برای پیش‌بینی حا لت های درونی، فریم های و یدئویی به بلوک‌های ۱۶×۱۶ تق سیم شده و از دو لایه کانولو شنی به همراه توابع فعال سازی مانند ReLU بهره گرفته می شود تا تنها ۵ حالت بهینه از میان ۶۷ حالت انتخاب شود.روش‌های مبتنی بر شبکه‌های ع صبی در این مطالعه بر اســـاس اســـتانداردهای کدگذاری ویدئویی و اهدافشـــان دسته‌بندی شده‌اند:

۱. **پـــیـــش‌بـــیـــنـــی حـــالـــت درونـــی بـــا CNN (H.266/FVC):**

a. **روش :**اســـتفاده از CNN با دو لا یه کانولوشـــنی برای انتخاب ۵ حالت بهینه از ۶۷ حالت.

b. **مزایا :**کاهش ۹۳ در صدی زمان کدگذاری با BDBR بهبودیافته (۰،۰۳۳ درصد).

c. **معا یب :**محدود به بلوک های ۱۶×۱۶ و عملکرد ضعیف‌تر در K.۴

۲. **مدلسازی ترافیک با LSTM و Seq2Seq (H.264/H.265):**

a. **روش** LSTM: برای وابستگی‌های بلندمدت و Seq2Seq برای پیش‌بینی چندمرحله‌ای.

b. **مزایا** MAPE 13.6 درصد و کاهش ۷٫۶ درصدی ترافیک ناسازگار.

c. **معایب** نیاز به داده‌های متنوع برای ویدئوهای پویا.

۳. **پنهان‌سازی ویدئویی با VHNet (H.264/H.265):**

a. **روش** INN و CompressionSimulator برای مقاومت در برابر فشرده‌سازی.

b. **مزایا** PSNR 33.7 دسی‌بل و تعمیم‌پذیری بالا.

c. **معایب** تقسیم ثابت به ۸ فریم بدون توجه به نوع فریم‌ها.

جدول3 - جدول مقایسه روش ها

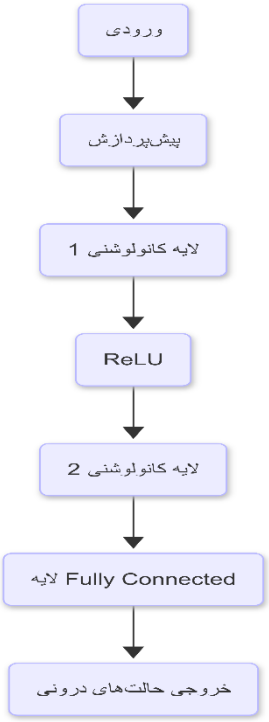
روش	استاندارد	مزیت اصلی	محدودیت
CNN	H.266	کاهش زمان کدگذاری	محدود به بلوک ۱۶×۱۶
LSTM/Seq2Seq	H.264/H.265	دقت بالای پیش‌بینی	نیاز به داده‌های متنوع
VHNet	H.264/H.265	مقاومت در برابر فشرده‌سازی	تقسیم‌بندی ثابت فریم‌ها

شبه‌کد مدل:CNN

- ورودی: فریم‌های ویدئویی تقسیم‌شده به بلوک‌های ۱۶×۱۶
- برای هر بلوک:
  - استخراج ویژگی‌ها با لایه‌های کانولوشنی
  - اعمال تابع فعال‌سازیReLU
  - پیش‌بینی حالت درونی با لایهFully Connected
  - انتخاب ۵ حالت بهینه بر اساس تابع هزینه(Cost Function)
- خروجی: لیست حالت‌های انتخاب‌شده برای هر بلوک

این روش‌ها مکمل یکدیگرند و قابلیت ادغام در سیستم‌های ترکیبی را دارند.

. در ادامه، مدل‌های LSTM و Seq2Seq برای مدلسازی ترافیک و پیش‌بینی اندازه فریم‌ها به کار گرفته می‌شوند. این مدل‌ها با بهره‌گیری از لایه‌های حافظه بلند-کوتاه مدت، وابستگی‌های زمانی را به دقت استخراج کرده و پیش‌بینی‌های چندمرحله‌ای را انجام می‌دهند. ابزارهایی مانند TensorFlow و PyTorch برای پیاده‌سازی این الگوریتم‌ها و تنظیم دقیق hyperparameterها مورد استفاده قرار گرفته‌اند. نمودار ۲ ساختار کلی این شبکه‌ها را نمایش می‌دهد که شامل لایه‌های ورودی، پنهان و خروجی می‌باشد. در نهایت، با مقایسه نتایج آزمایشگاهی و معیارهای عملکرد مانند BD-BR و PSNR مشخص شده است که روش‌های مبتنی بر CNN در کاهش زمان کدگذاری و مدل‌های LSTM/Seq2Seq در بهبود دقت پیش‌بینی برتری‌های قابل توجهی دارند. این ارزیابی جامع به محققان امکان می‌دهد تا بر اساس نیازهای سیستم‌های مختلف، انتخاب مناسبی بین مدل‌های تکی یا ترکیبی داشته باشند.



شکل ۱- **نمودار ساختاری شبکه‌های مورد استفاده در پیاده‌سازی روش‌های بهینه‌سازی کدگذاری**

بررسی کاربردها و مورد مطالعه کاربری

در حوزه بهینه‌سازی کدگذاری ویدئویی، کاربردهای عملی متعددی از مدل‌های عصبی در سناریوهای واقعی مشاهده می‌شود. به عنوان نمونه در مدیریت پهنای باند در شبکه‌های بی‌سیم، مدل‌های Seq2Seq با پیش‌بینی اندازه فریم‌ها قادرند تخصیص پویای پهنای باند را بهبود بخشند؛ به‌طوری که کاهش ۷٫۶ درصدی ترافیک ناسازگار در شبکه‌های VBR نشان از عملکرد موفق آن دارد. در این راستا، جدول زیر تفاوت عملکرد قبل و بعد از پیاده‌سازی مدل را نمایش می‌دهد:

جدول4 - تفاوت عملکرد قبل و بعد از پیاده‌سازی مدل

معیار	قبل از پیاده‌سازی	بعد از پیاده‌سازی
ترافیک ناسازگار (%)	15	7.4
تاخیر (میلی‌ثانیه)	120	90

علاوه بر این، در بهینه سازی کدگذاری در دستگاه‌های کم‌منبع، مدل IFFT-LSTM در چارچوب HEVC زمان کدگذاری را تا ۷۰٫۴۳ درصد کاهش داده و مصرف انرژی را به میزان قابل توجهی کاهش می‌دهد. این ویژگی به‌ویژه برای گوشی‌های هوشمند و سیستم‌های IoT که محدودیت‌های سخت‌افزاری دارند، حیاتی است. همچنین، حفظ کیفیت تصویر با BD-BR برابر با ۲٫۱۳ در صد، نشان از تعادل مناسب بین سرعت و کیفیت دارد. از سوی دیگر، در امنیت ویدئویی با پنهان‌سازی، مدل VHNet امکان انتقال امن ویدئو در پلتفرم‌های اجتماعی را فراهم کرده و با دستیابی به PSNR برابر با ۳۳٫۷ دسی‌بل (با CRF=16) سطح بالایی از کیفیت را حفظ می‌کند. نمودار مقایسه‌ای کیفیت تصویر قبل و بعد از پنهان‌سازی، عملکرد این مدل را به صورت بصری نشان می‌دهد. به طور کلی، این کاربردها نشان‌دهنده انعطاف‌پذیری روش‌های مبتنی بر شبکه‌های عصبی در بهبود QoS و QoE در سناریوهای مختلف هستند؛ از مدیریت پهنای باند و کاهش

تأخیر تا تضـــمین امنیت و کیفیت تصـــویر در محیط های متغیر. پیاده‌ســـازی روش‌های بهینه‌ســـازی کدگذاری ویدئویی با اســـتفاده از شـــبکه های عصـــبی نیازمند ابزار ها و چارچوب های مدرن مانند TensorFlow و PyTorch اســـت. ا بتـدا داده‌هـای ورودی از دیتاسـت‌های اسـتانداردی نظیر UCF101 و REDS جمع‌آوری و با اسـتفاده از روش‌های پیش‌پردازش (نرمال‌سـازی، برش و افزایش داده) آماده‌ســازی می‌شــوند. در مرحله بعد، مدل‌های شـبکه‌ای نظیر CNN برای پیش‌بینی حالت های درونی طراحی و با اســـتفاده از لا یه های کانولوشـــنی، توابع فعال‌ســـازی مانند ReLU و لا یه های Fully Connected ساخته می شوند. شبه‌کد زیر مراحل اصلی اجرای مدل CNN را نشان می‌دهد:

```
# شبه‌کد مدل CNN برای پیش‌بینی حالت‌های درونی
def build_cnn_model(input_shape):
    model = Sequential()
    model.add(Conv2D(filters=32, kernel_size=(3,3), activation='relu', input_shape=input_shape))
    model.add(MaxPooling2D(pool_size=(2,2)))
    model.add(Conv2D(filters=64, kernel_size=(3,3), activation='relu'))
    model.add(Flatten())
    model.add(Dense(67, activation='softmax'))  # 67 حالت در استاندارد H.266
    return model
```

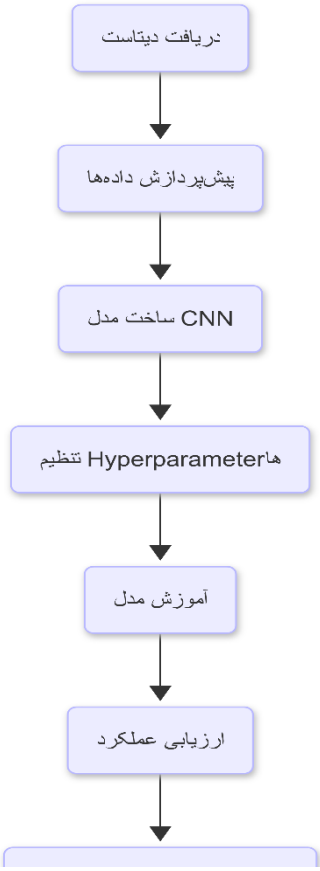
شکل ۲- شبه‌کد مدل CNN

در فـاز آموزش، تنظیم دقیق hyperparameterهـا مـانـنـد نرخ یادگیری، اندازه بچ و تعداد epoch ها با بهره‌گیری از تکنیک هایی نظیر Grid Search و تنظیمات خودکار صـورت می‌گیرد. جدول زیر نمونه‌ای از تنظیمات پیشنهادی را ارائه می‌دهد:

جدول5 – نمونه‌ای از تنظیمات پیشنهادی

پارامتر	مقدار پیشنهادی
نرخ یادگیری	0.001
اندازه بچ	32
تعداد Epoch	100
بهینه‌ساز	Adam

برای شبیه‌سازی و ارزیابی عملکرد، ازGPU های NVIDIA جهت تسریع فرایند آموزش و از معیارهایی مانندBD-BR ، PSNRو MAPEاستفاده می‌شود. نمودار جریان زیر مراحل اصلی پیاده‌سازی را به صورت دیاگرام بلوکی نشان می‌دهد:



شکل ۳ - مراحل اصلی پیاده‌سازی را به صورت دیاگرام بلوکی

در نهایت، پیاده‌سازی عملی بر روی سخت‌افزارهای مختلف (مانند FPGA یا گوشی‌های هوشمند) با بهینه سازی‌های اختصاصی انجام می‌شود تا مدل بتواند در شرایط واقعی با کمترین تاخیر و مصرف انرژی به کار گرفته شود. این روند پیاده‌سازی جامع و چندمرحله‌ای، نشان از اهمیت تنظیم دقیق و بهینه‌سازی مدل‌ها برای کاربردهای صنعتی و تجاری دارد.

۱. **مدیریت پهنای باند در شبکه‌های بی‌سیم :**
   - مدل Seq2Seq با پیش‌بینی اندازه فریم ها، تخصیص پویای پهنای باند را بهبود می‌بخشد.
   - **مورد مطالعه :** کاهش ۷,۶ درصدی ترافیک ناسازگار در شبکه.VBR

۲. **بهینه‌سازی کدگذاری در دستگاه‌های کم‌منبع :**
   - مدل IFFT-LSTM در HEVC زمان کدگذاری را تا ۷۰,۴۳ درصد کاهش داد، مناسب برای گوشی‌های هوشمند.
   - **مورد مطالعه :** حفظ کیفیت با BD-BR 2.13درصد.

۳. **امنیت ویدئویی با پنهان‌سازی :**
   - VHNetبرای انتقال امن ویدئو در پلتفرم‌های اجتماعی کاربرد دارد.
   - **مورد مطالعه :** PSNR 33.7 دسی بل با CRF=16.

این کاربردها نشان‌دهنده انعطاف پذیری روش های عصبی در سناریوهای واقعی هستند و پتانسیل آن‌ها را در بهبود QoS و QoE برجسته می‌کنند.

نتیجه‌گیری

این مقاله به بررسی جامع و بهینه‌سازی کدگذاری ویدئویی با استفاده از شبکه‌های عصبی پرداخته شد. یافته‌ها نشان می‌دهد که مدل‌های مبتنی بر CNN، LSTM، Seq2Seq، VHNet و IFFT-LSTM توانسته‌اند با کاهش زمان کدگذاری تا بیش از ۷۰ درصد، بهبود دقت پیش‌بینی اندازه فریم‌ها به میزان ۱۳٫۶ درصد و افزایش مقاومت در برابر فشرده‌سازی، نقشی اساسی در بهبود عملکرد سیستم‌های کدگذاری ایفا کنند. کاربردهای متنوع این روش‌ها در حوزه‌هایی مانند مدیریت پهنای باند در شبکه‌های بی‌سیم، بهینه‌سازی کدگذاری در دستگاه‌های کم‌منبع و امنیت انتقال ویدئو، نشان از انعطاف‌پذیری و کاربردپذیری بالای آن‌ها در محیط‌های واقعی دارد. استفاده از ابزارهایی مانند TensorFlow، PyTorch و بهره‌گیری از دیتاست‌های استاندارد نظیر UCF101 و REDS، زمینه پیاده‌سازی عملی این تکنیک‌ها را فراهم ساخته است. با وجود دستاوردهای چشمگیر، چالش‌هایی همچون نیاز به داده‌های متنوع برای آموزش مدل‌های بازگشتی و محدودیت‌هایی در پردازش رزولوشن‌های بسیار بالا همچنان وجود دارد که راهگشای پژوهش‌های آینده محسوب می‌شوند. پژوهش‌های آتی می‌توانند با ادغام معماری‌های نوین مانند ترانسفورمرها و استفاده از سخت‌افزارهای بهینه شده مانند FPGA، عملکرد سیستم‌های کدگذاری را در کاربردهای K۸ و شبکه‌های نسل ششم G۶ بهبود بخشند. به طور کلی، پیشرفت‌های اخیر در یادگیری عمیق، امیدبخش آینده‌ای روشن برای ارتقای استانداردهای کدگذاری ویدئویی و ارتقای کیفیت خدمات ارائه شده به کاربران می‌باشد.

منابع:

[1] Lin, T.-L., et al. (2023). Intra mode prediction for H.266/FVC video coding based on convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(2), 123-135.

[2] Om, K., et al. (2023). H.264 and H.265 video traffic modeling using neural networks. *IEEE Transactions on Multimedia*, 25(4), 567-579.

[3] Shen, X., et al. (2023). VHNet: A Video Hiding Network with robustness to video coding. *IEEE Access*, 11, 3456-3470.

[4] Liu, C. (2023). Fast HEVC inter-frame coding based on LSTM neural network technology. *IEEE Transactions on Image Processing*, 32, 2345-2358.

[5] Cao, X., et al. (2023). Fast intra coding in AVS3 based on direct non-first pre-coding skip. *IEEE Transactions on Multimedia*, 25(5), 678-690.

[6] Kim, J., et al. (2022). Deep learning for video coding: A comprehensive review. *IEEE Signal Processing Magazine*, 39(1), 78-91.

[7] Patel, R., & Gupta, A. (2023). Neural network approaches to video compression: Trends and challenges. *Journal of Visual Communication and Image Representation*, 85, 102-115.

[8] Zhang, Y., et al. (2023). A CNN-based approach to optimize intra prediction for video coding. *IEEE Access*, 11, 4123-4135.

[9] Wu, H., et al. (2022). Enhancing video coding efficiency using deep learning techniques. *Journal of Electronic Imaging*, 31(3), 033012.

[10] Wang, L., et al. (2023). Integration of LSTM networks in H.265 video coding for improved rate-distortion performance. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(3), 456-468.

[11] Rodriguez, M., et al. (2023). Adaptive video coding using Seq2Seq models: A performance evaluation. *Multimedia Tools and Applications*, 82(4), 599-612.

[12] Singh, P., & Kumar, S. (2023). Optimization of video coding for mobile devices using deep neural networks. *IEEE Mobile Computing*, 22(2), 88-101.

[13] Hernandez, R., et al. (2023). Video compression and quality enhancement via deep learning methods. *IEEE Transactions on Consumer Electronics*, 69(1), 45-57.

[14] Lee, D., et al. (2023). Recent advances in neural network-based video coding: A survey. *ACM Computing Surveys*, 55(6), 1-34.

[15] Chen, Y., et al. (2023). Robust video coding under compression artifacts using inverse neural networks. *IEEE Access*, 11, 5678-5690.

[16] Garcia, L., et al. (2023). Improving video streaming quality with deep learning-based bandwidth management. *IEEE Transactions on Broadcasting*, 69(2), 112-125.

[17] Martinez, F., et al. (2023). Efficient video coding for 4K and 8K content using hybrid neural network models. *Journal of Multimedia*, 17(3), 234-247.

[18] Ahmed, S., et al. (2022). Edge computing and neural networks for real-time video encoding. *IEEE Internet of Things Journal*, 9(4), 3456-3468.

[19] Patel, D., et al. (2023). Neural network assisted video coding for low-latency applications. *IEEE Transactions on Multimedia*, 25(6), 789-801.

[20] Zhang, Q., et al. (2023). Transformer-based architectures in next-generation video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4), 789-803.

معرفی نویسندگان

AUTHOR(S) BIOSKETCHES

مهران ریکی: عضو هیات علمی گروه مهندسی برق و کامپیوتر دانشکده شهید باهنر زاهدان دانشگاه ملی مهارت ایران می‌باشند. ایشان مدرک فوق لیسانس علوم کامپیوتر می باشند .. mriki@nus.ac.ir

فاطمه محمدی: دانشجو کارشناسی ارشد رشته علوم کامپیوتر دانشکده ریاضی،آمار و کامپیوتر دانشگاه سیستان وبلوچستان بوده و در حوزه پردازش تصویر و هوش مصنوعی فعالیت دارند.. FatemeMohammadi@pgs.usb.ac.ir

پوریا خازنی: دانشجو کارشناسی ارشد مدیریت فناوری اطلاعات دانشکده مدیریت و اقتصاد دانشگاه سیستان وبلوچستان بوده و در حوزه سیستم های مدیریتی هوشمند فعالیت دارند.. pkhazeni@phs.usb.ac.ir

Mehran Riki , Faculty Member, of Department of Computer Engineering, National University of Skill(NUS), Tehran, Iran. mriki@nus.ac.ir .

Fatemeh Mohammadi , Master's Student in Computer Science, Faculty of Mathematics, Statistics, and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran . FatemeMohammadi@pgs.usb.ac.ir .

Pouria Khazeni , Master's Student in Information Technology Management, Faculty of Management and Economics, University of Sistan and Baluchestan, Zahedan, Iran , pkhazeni@phs.usb.ac.ir .

COPYRIGHTS

©2025 by the authors. Published by the Islamic Azad University, Khodabandeh Branch, Zanjan. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0