|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 重庆邮电大学  CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS | | | | |
| 博士学位论文  DOCTORAL DISSERTATION | | | | |
|  | | | | |
|  | | | | |
| **论文题目** | | **基于深度神经网络的HEVC帧内快速** | |  |
|  | | **深度决策算法研究** | |  |
|  | |  | |  |
|  | **学科专业** | | **计算机科学与技术** |  |
|  | **学 　号** | | **S210231282** |  |
|  | **作者姓名** | | **周广义** |  |
|  | **指导教师** | |  |  |
|  | **学　　院** | |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学校代码 | 10617 | | | | | UDC |  | |
| 分类号 |  | | | | | 密级 |  | |
| 学　位　论　文 | | | | | | | | |
| **基于深度神经网络的HEVC帧内快速深度决策算法研究** | | | | | | | | |
|  | | | | | | | | |
|  | | | | **周广义** | | |  | |
|  | | | |  | | |  | |
|  | | | | | | | | |
| 指导教师 | | **某某某** | | | | | **教　授** |  |
|  | | **某 某** | | | | | **副教授** |  |
|  | |  | | | | |  |  |
|  | |  | | | | |  |  |
|  | |  | | | | | |  |
|  | |  | | | | | |  |
|  | |  | | | | | |  |
| 申请学位级别 | | **硕士** | | | 学科专业 | |  | |
| 专业学位领域 | |  | | | | | | |
| 答辩委员会主席 | | **某某某 教 授** | | | 论文答辩日期 | | **2021年5月20日** | |
| 学位授予单位和日期 | | | **重庆邮电大学** | | | | **2021年6月** | |

|  |  |
| --- | --- |
| **Intra-frame fast depthdecision algorithm for HEVC based on deep neural network** | |
| A Doctoral Dissertation Submitted to  Chongqing University of Posts and Telecommunications | |
|  | |
| Discipline | **Computer Science and Technology** |
| Student ID | **S210231282** |
| Author |  |
| Supervisor |  |
| School | **School of Computer Science and Technology** |

**重庆邮电大学**

**学位论文独创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文中不包含其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

**重庆邮电大学**

**学位论文使用授权书**

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于 ：

□**公开论文**

□**涉密论文**，保密 年，过保密期后适用本授权书。

（请在以上方框内选择打“**√**”）

作者签名： 导师签名：

日期： 年 月 日

**摘　要**

随着大数据时代的到来和无时不在的移动互联网通信技术的发展，人们通信方式的内容形式已经变成了大量的图片和视频。为了方便视频数据的保存与传输最有效的方法是将视频进行压缩处理，现阶段最流行也是最普遍的处理视频压缩的方式是（高效视频编码HEVC），HEVC在上一代视频压缩标准（Advanced Video Coding：AVC）的基础上提高了50%的压缩率，同时带来了编码复杂度的提升，在当前的高效视频编码（HEVC）领域，帧内编码树单元（CTU）的深度划分决策对编码效率和视频质量具有显著影响。然而，传统的深度划分算法往往依赖于复杂的启发式规则或简单的机器学习方法，这限制了划分性能的提升并增加了编码复杂度。为了解决这些问题，本文旨在来研究通过利用深度神经网络（DNN）的强大特征提取和分类能力，实现更高效且更准确的CTU深度划分。从而降低HEVC的编码复杂度，本文的研究工作如下

（1）本研究的第一个贡献是设计了一种多尺度多输入的卷积神经网络（MMCNN），它能够同时处理多个不同量化参数（QP值）下的图像块输入，并且考虑不同尺度的CU特征。这种设计使得网络能够捕捉到跨尺度的特征关联，从而提升了特征的表达能力并增强了网络对不同QP条件下的适应性。

（2）通过分析研究工作（1）中提出方法的一些局限性，针对分类结果错误累积问题，本文提出了一种互补分类策略（CCS）。该策略由两个独立的(Multi-scale-multi-input Complementation Classification Network :MMCNN)组成，分别对CTU中的32×32 CU进行三分类深度预测，以及对16×16 CU进行二分类划分预测。通过将两个网络的预测结果进行融合，并采用投票机制来确定每个CU的最终深度值，实现了分类结果的互补，从而显著提高了深度划分的准确性。

为了验证所提出方法的有效性，我们在多个HEVC标准测试序列上进行了广泛的实验。实验结果表明，与现有的帧内深度划分方法相比，我们的MCCN框架在保持相似编码效率的同时，可以有效降低编码复杂度。此外，所提出的方法具有较好的泛化能力，即在不同类型的视频内容上都能达到稳定的性能提升。本文提出的模型在客观评价指标和主观视觉效果两个方面都获得了良好的效果

**关键词：**高效视频编码，编码树单元，多尺度多输入，深度神经网络，互补分类网络

**ABSTRACT**

Dissertation /Thesis is postgraduate’s main academic performance to display her/his works of scientific research, which shows the author’s new invention, new theory or new opinion in her/his research. It is the crucial document for the graduate students to apply for degree, and it is also the important scientific research literature and the valuable wealth of society.

In order to further standardize the format of dissertation/thesis writing and improve graduate dissertation/thesis quality, this temolate is formulated with reference to the national standard "Rules for Dissertation Writing" (GB/T 7713.1-2006) and the reality of CQUPT.

**Keywords:**Dissertation/Thesis, Writing Specification,Thesis Template, Chongqing University of Posts and Telecommunications

目　录

[摘　要 I](#_Toc153361582)

[ABSTRACT II](#_Toc153361583)

[图目录 V](#_Toc153361584)

[表目录 VI](#_Toc153361585)

[主要符号表 VII](#_Toc153361586)

[缩略词表 VIII](#_Toc153361587)

[第1章 绪论 1](#_Toc153361588)

[1.1 研究背景及意义 1](#_Toc153361589)

[1.2 国内外研究现状 1](#_Toc153361590)

[1.3 论文研究主要内容 1](#_Toc153361591)

[1.4 论文组织结构 1](#_Toc153361592)

[第一章 绪论 2](#_Toc153361593)

[（一）研究背景及意义 2](#_Toc153361594)

[1. 研究背景 2](#_Toc153361595)

[2. 研究意义 2](#_Toc153361596)

[（二）国内外研究现状 2](#_Toc153361597)

[（三）论文研究主要内容 2](#_Toc153361598)

[（四）论文组织结构 2](#_Toc153361599)

[第2章 论文结构及文字格式 3](#_Toc153361600)

[2.1 本章引言 3](#_Toc153361602)

[2.2 论文结构 3](#_Toc153361603)

[2.3 字数要求 3](#_Toc153361604)

[2.3.1 硕士论文要求 3](#_Toc153361605)

[2.3.1 博士论文要求 4](#_Toc153361606)

[2.4 字体和段落 4](#_Toc153361607)

[2.5 本章小结 5](#_Toc153361608)

[第3章 图表、公式格式和印制要求 6](#_Toc153361609)

[3.1 本章引言 6](#_Toc153361611)

[3.2 图和表格式 6](#_Toc153361612)

[3.2.1 图 6](#_Toc153361613)

[3.2.2 表 8](#_Toc153361614)

[3.3 公式格式 9](#_Toc153361615)

[3.4 印制要求 10](#_Toc153361616)

[3.5 本章小结 10](#_Toc153361617)

[第4章 总结与展望 11](#_Toc153361618)

[4.1 主要结论 11](#_Toc153361620)

[4.2 研究展望 11](#_Toc153361621)

[参考文献 12](#_Toc153361622)

[附录A 各学院中英文名称对照表 15](#_Toc153361623)

[作者简介 16](#_Toc153361624)

[1. 基本情况 16](#_Toc153361625)

[2. 教育和工作经历 16](#_Toc153361626)

[3. 攻读学位期间的研究成果 16](#_Toc153361627)

[3.1 发表的学术论文和著作 16](#_Toc153361628)

[3.2 申请（授权）专利 16](#_Toc153361629)

[3.3 参与的科研项目及获奖 16](#_Toc153361630)

[致 谢 17](#_Toc153361631)

图目录

[图2‑1 学位论文基本结构 3](#_Toc152868291)

[图3‑1 不同缩放系数 ν 的缩放结果 6](#_Toc152868292)

[图3‑2 HARP模型中通过边坍塌进行结构粒度粗化的策略[64] 。 7](#_Toc152868293)

[图3‑3 HARP模型中通过边坍塌进行结构粒度粗化的策略[64] 7](#_Toc152868294)

表目录

[表2‑1 中、英文字号对应关系 4](#_Toc128917515)

[表2‑2 主要文字及段落格式要求 4](#_Toc128917516)

[表3‑1 电流类型对效率的影响 7](#_Toc128917517)

[表3‑2 球队的比赛结果统计表[5] 8](#_Toc128917518)

[表3‑3 学位论文页面设置 9](#_Toc128917519)

主要符号表

|  |  |  |
| --- | --- | --- |
| **符号** | **说明** | **页码** |
| 𝒫Ω(·) | 集合Ω上的投影算子 | 6 |
| *c* | 电磁波的相平面速度 | 10 |

缩略词表

|  |  |  |
| --- | --- | --- |
| **英文缩写** | **英文全称** | **中文全称** |
| CQUPT | Chongqing University of Posts and Telecommunications | 重庆邮电大学 |
| IEEE | Institute of Electrical and Electronics Engineers | 电气和电子工程师协会 |

第一章 绪论

## 1.1研究背景及意义

随着大数据时代的到来和无时不在的移动互联网通信技术的发展，以及互联网、物联网的全面覆盖，一个以大数据、大流通为特征的信息时代已经开启。人们对于具有生动、全面、信息量大为特征的视频信息的需求日益增大，使得大数据中的海量视频信息的占比达到一个新高度。随着4K、8K等高清视频格式的普及，视频数据量急剧增加，对存储和带宽提出了更高的要求，移动设备和社交媒体的普及导致视频内容的生产和消费量大幅上升。传统的视频编码标准（如H.264）在高清视频传输和存储方面面临带宽和存储限制。 网络服务提供商和内容分发网络寻求降低数据传输成本。因此视频压缩变得越来越重要。

新一代的视频压缩标准（High Efficiency Video Coding，HEVC）的出现，旨在提供比H.264更高的数据压缩率，理论上能够在保持相同视频质量的情况下将码率降低约50%。这意味着相同质量的视频，使用HEVC编码后，文件大小和所需的传输带宽大大减少。HEVC的出现使得更复杂的编码算法成为可能，它提出了编码树单元（Coding Tree Unit：CTU）的概念，每一个CTU可进行不同深度的划分，从64×64到最小的8×8大小不等，同时还将预测模式提高到了35种之多，以提高预测块的预测精度。通过递归地计算所有深度上所有图像块的率失真代价（Rate-Distortion cost：RD cost）来决定每个编码单元（Coding Unit ：CU)的深度。从而达到最优的编码效果，由于这个过程，计算复杂度和编码时间显着增加，HEVC 的编码时间比（H.264/AVC）平均高253%。这导致HEVC的编码复杂度变得极大，尤其是针对CTU深度的划分[1]-[4]。因此，如何能够在保证较低BD-BR（Bjøntegaard delta bit rate）的同时尽可能提升CTU深度划分的效率降低HEVC的编码复杂度已经成为了近年来人们研究的热点。

为了降低HEVC的编码复杂度，减少编码时间，人们提出了两大类提升CTU深度划分效率的方法：基于启发式的方法（Heuristic-based Methods）和基于深度学习的方法（Deep-learning-based Methods）。对于启发式的方法而言，几乎都是通过人为提取图像中的纹理信息特征或者人为设置某些规则来提前终止CTU的深度划分，以避免暴力递归的RDO（Rate Distortion Optimization）搜索，从而达到降低深度划分的复杂度的目的，例如：利用CTU中每个CU的纹理特征以及其与相邻CU的空间相关性来决定CTU的划分结果[5]-[18]。

近年来，随着人工智能的发展。深度学习神经网络，尤其是卷积神经网络（Convolutional Neural Network：CNN）已经被广泛用于优化HEVC帧内模式的CTU深度划分中。用神经网络代替RDO的计算过程，通过给网络输入当前编码块的像素值，使用神经网络去做出深度的划分，这将大大节省HEVC的编码时间。

但是，无论是基于启发式的方法还是基于学习的方法，它们的性能仍具有较大的提升空间，网络模型需要进一步提升，划分策略也需进一步完善。因此，综上所述，本文的研究目标是充分利用神经网络在图像特征提取上的优势与HEVC编码系统相结合，进一步降低HEVC系统的编码复杂度。

## 1.2国内外研究现状

由HECV高压缩率所带来的复杂度的提升，一直是近年来的研究热点问题，在帧内编码过程中，发现CU划分的递归过程贡献了最大比例的编码时间，大约占据整个编码时间的80%。由于CU的深度划分是影响编码时间的重要部分。因此，在过去的五年中，出现了大量降低HEVC编码复杂性的方法。在这些HEVC复杂性降低方法中，大多数都试图简化CU分区的过程。这些方法的基本思想是提前预测CU分区，而不是蛮力递归RD优化 (RDO) 搜索。

预测CU分区的早期工作是启发式方法，它们探索一些中间特征，以便在检查所有可能的之前尽早确定CU划分。在启发式的方法中，也出现了一些基于机器学习的方法，但它仍需要人工设置一些CTU纹理和与相邻CTU之间的空间相关特征来描述CTU的分区。

在近些年，旨在加速HEVC帧内编码的研究一直在从基于启发式的策略转变为基于学习的策略，从而出现了大量基于深度学习的方法（Deep-learning-based Methods）。在这些方法中，大量使用了CNN（Convolutional Neural Network）网络结构，来预测CTU的划分。

### 基于启发式的方法

为了有效降低HEVC帧内编码的复杂度，人们提出了大量的基于启发式的方法来提升CTU深度划分效率。基于启发式的方法通常是通过人为提取CTU中CU的特征或者人为设置某些规则来提前终止CTU的深度划分过程，以避免暴力递归的RDO搜索，从而达到降低深度划分复杂度的目的。空间/时间/深度相关性、统计率失真 (RD) 和纹理信息是预测CU深度决策的三个常用特征。

在Chen[5]等人的研究中就利用了时间相关性来预测CU的深度范围，具体来说基于对前一帧获得的模式信息的分析来衡量视频内容的复杂度。然后，基于该特征与编码树单元（CTU）深度范围的关系构建模型。根据模型，跳过了编码单元 （CU） 划分的不必要操作。Cen[6]等人提出一种利用序列帧空间相关性的快速编码单元（CU）深度决策机制。所提出的机制包括一种自适应CU深度范围确定和CU深度比较算法。CU深度范围根据CU深度在同一序列中的分布确定。因此，可以跳过范围之外的 CU深度的速率失真（RD）成本计算。此外，通过将相邻 CU 的深度与当前CU深度进行比较，可以进一步跳过当前CU深度的RD成本计算。

Xiong[7]等人通过研究运动发散与率失真（RD）成本之间的隐含关系，提出了一种基于金字塔运动发散的模型来快速决定CU是否需要划分。Kim[8]等人]提出了一种根据每个CU中高频关键点的数量来决定CU是否分裂的方法。Shen[9]等人提出通过使用重要且计算友好的特征来做出CU大小决策，并定义了贝叶斯决策规则，以通过同时最小化贝叶斯风险来帮助准确快速地选择CU大小。

此外，基于编码图像的时间和空间特性，Wang[10]等人介绍了一种基于纹理复杂度和运动特性的剪枝策略。然后，进一步使用相邻编码块的运动信息来快速确定参考图像，从而减少编码时间。Liu[11]等人提出了一种基于纹理复杂度和图像方向的预测模式自适应决策方法。

Shen[12]等人提出了一种用于HEVC内编码的快速CU大小决策算法，通过减少每个CTU中需要检查的候选CU数量来加快该过程。该算法的新颖性在于（1）基于纹理均匀性，提出了一种基于自适应阈值的早期CU大小决策确定方法；（2）基于纹理属性和相邻编码CU的编码信息的组合，提出了一种新的大CU内部预测旁路策略。在文献[13]中，Gweon等人通过在亮度和色度的所有编码块标志Coded Block Flags（CBF）为零的情况下跳过对当前CU深度的RDcost计算来减少计算复杂度。

Choi[14]等人通过判断当前CU深度下的最佳预测模式是否为SKIP来提前结束对该CU深度的划分。Yang[15]等人通过采用差分运动矢量和编码块标志两个特征作为CU是否划分的标准来降低编码器的复杂度。Cho[16]等人使用了具有低复杂度和低RDcost的贝叶斯决策规则来对CU的划分做出判决。Zhang[17]等人采用时空相关性进行自适应的CU深度搜索，从而降低CTU划分的复杂度。Zhao[18]等人采用低成本的Hadamard cost来作为终止CU划分的标准。Khan[19]等人提出了一种基于视频帧内容自适应的快速CU决策方法：根据视频帧中不同局部区域的内容，自适应的将较小的CU组合成较大的CU。

这些传统的方法在一定程度上提高了编码单元分割的效率，降低了HEVC的编码复杂度。在之后出现了一些使用机器学习的方法来进一步挖掘图像的纹理特征。

在文献[20]中，Zhang等人提出了一种基于梯度的机器学习粗模式决策和RD优化算法。为了实现快速的CU划分决策，Zhang等人提出了两个以深度差和RD代价为特征的线性支持向量机来进行早期CU分裂和早期CU终止决策。

Hu[21]等人基于离散余弦变换系数透明复合模型，将帧内编码模式判决问题转换成贝叶斯判决问题，提出了一种基于离群点的快速帧内模式判决算法。Liu[22]等人提出了一种基于支持向量机的三分类器来判断CU是否需要划分。Jamali[23]等人提出了一种针对帧内模式的RDO成本预测方法和一种基于梯度的模式筛选方法。Jamali等人采用Prewitt算子来从候选模式列表中筛选出方向相关的模式。由以上分析可知，现有的基于启发式的方法几乎都是通过人为手动的对图像中某些成分的特征（如边缘、纹理）来设置一些规则，使得满足这些特征的CU可以提前终止划分，来达到降低HEVC帧内编码复杂度的目的。然而自然图像中边缘和纹理是千变万化的，不同CU之间的边缘/纹理特征以及它们之间的关系是十分复杂的，这些算法不能正确地考虑在各种序列中的CTU的分割趋势。也就是说，它们没有根据图像纹理和编码信息综合考虑CTU划分趋势。它们使用在编码过程中可以获得的各种因素来决定是否跳过对RD成本的操作。然而，存在太多的因素，并且难以具体地确定哪些与CTU划分相关。因此，仅参考元素中的一些来确定CTU分割，这可能导致较差的预测。这种人为设置相关规则的方式是几乎不可能满足所有的CU的。因此，虽然启发式的方法能够在一定程度上降低HEVC的编码复杂度，但其性能仍有较大提升空间。

### 基于学习的方法

近年来，随着深度学习神经网络的出现，尤其CNN，因为其强大的特征提取能力和非线性表达能力和自动特征学习，它们可以有效地用于各种视频编解码场景，从而增强了所提出的算法的通用性，因此，基于CNN的方法越来越受到关注，并在视频编码和解码领域中找到各种应用，目前已经被广泛用于优化HEVC帧内模式的CTU深度划分中。通过神经网络可以很好的代替人工方式来提取CTU中的图像特征和纹理信息，来提前终止CU的深度划分。

Yu[24]等人提出了一种基于CNN的二分类CTU深度划分方法：从初始的64×64图像块开始，判断其是否需要划分，如需要划分则继续对下一层的每个32×32图像块进行二分类判断，如不需要划分则直接终止对该层CU的划分，他们在非常有限的环境下应用CNN。编码仅在INTRA模式下执行，该模式具有相对较轻的计算负载。此外，他们为所有深度设计了相同的神经网络架构，无法有效地考虑每个CU深度划分的属性。

在文献[25]中，Li等人使用CNN来降低INTRA模式下的CTU决策复杂度，他们针对不同深度的CU设计了不同的神经网络以实现更具针对性的CU深度预测，但是Li等人却忽略了不同尺寸CU间的关联性。因此它不能很好地反映CTU分区趋势。Kim[26]设计了一种基于神经网络推理的编码单元深度判决算法，该算法不仅适用于帧内模式，而且适用于帧间模式，并根据神经网络推理的结果将编码单元划分为64×64、32×32、16×16三种深度。此外，通过为四个量化参数（QP）设计每个网络，可以适用于各种QP。Xu[27]等人提出了一种ETH-CNN（Early Terminated Hierarchical-CNN）来进行分等级的CU深度地图预测，以降低帧内模式HEVC的复杂度。Kuanar[28]等人提出了一种基于卷积神经网络的算法，该算法可以学习区域图像特征并执行分类工作将空间模式分为四类，这些分类结果用于编码器下游系统，以在每个CTU块中寻找最佳编码单元，从而减少预测模式的数量。

Li[29]等人提出了一种基于CNN的速率失真优化算法，通过结合混合注意力机制模块，以优化HEVC中CTU划分的RDO。首先，Li等人设计了一个与当前CU大小相兼容的CNN来预测CU的深度划分。其次，他们设计了一个卷积块，以增强CU块之间的信息交互，并将卷积块注意力模块（Convolution Block Attention Module:CBAM）引入到CNN中，构建了CNN-CBAM。该模块能够正确关注目标对象，提高CU划分的准确率。

针对CU深度划分的复杂递归计算，Qin[30]等人同样提出了一种基于CNN的CU划分方法：设计了一种改进的VGGNet来替代暴力搜索策略，有效降低了帧内编码的计算复杂度。Fan[31]等人提出了一种采用块划分神经网络来预测CTU的块划分结构的方法。Fan等人采用自适应阈值技术来准确管理CNN的预测误差。Wang[32]等人提出了一种基于密集连接卷积神经网络（Densely Connected Convolution Neural Network:D-CNN）来的CU深度划分方法。Wang等人首先设计了一个密集连接块来充分提取CTU的像素特征；然后，Wang等人采用有效的通道注意力（Efficient Channel Attention: ECA）和自适应卷积核大小应来捕获D-CNN卷积通道的特征信息；最后，Wang等人制定了一种阈值优化策略，通过为每个深度选择一个最佳阈值来平衡视频编码的计算复杂性和RD性能。Feng[33]等人提出了一种面向HEVC帧内编码的CNN快速CTU深度划分方法，但是该方法仅以单个尺度的CTU作为输入，导致网络难以学习到小尺寸CU的特征。Hari[34]等人采用不同大小的卷积核（5×5、4×4、2×2）来预测CTU的深度地图。

Lorkiewicz等人[35]提出了一种预训练的神经网络（Artificial Neural Network：ANN）来预测CTU的深度地图，但Lorkiewicz等人同样只考虑了单一尺度的CTU输入。

Ren[36]等人提出了一种基于CNN的帧内块划分（intra block partition CNN：IBP-CNN）网络，通过学习图像块的纹理特征来实现块的深度划分。Feng等人[37]提出了一种基于CNN的HEVC帧内深度范围（HEVC intra depth range CNN：HIDR-CNN）预测网络，通过预测CTU可能的深度范围，并仅在预测的深度范围内执行RDcost计算来减少RDcost的计算次数，从而达到降低编码复杂度的目的。

Li[38]等人提出了一种端到端的快速算法，以辅助帧内编码中CTU的结构划分决策。Imen[39]等人基于LeNet-5和AlexNet，有效降低了检查所有块决策候选者的计算时间，从而深度优化了HEVC编码单元划分模块的全帧内配置。Yao等人[40]构建了一种基于CNN的双网络模型：预测网络和目标网络。通过基于RD函数的强化学习方法，在确定最优CU划分模式的同时获得CU分区的最佳RD估计。

由以上分析可知，虽然现有的基于CNN的深度预测方法能够通过自主学习CTU的结构来做出对应的深度划分决策，但是，一方面，这些方法几乎都是仅采用了一种串联的二分类策略来决定CTU的划分结果，存在着明显的分类错误累积影响。另一方面，现有的基于CNN的CTU深度划分方法几乎都只考虑了单一尺度的输入，而忽略了不同尺度CU之间的特征关联。

## 论文研究的主要内容

在众多的针对加快HEVC编码速度、降低其编码复杂度的方法中本论文关注和研究的重点基于深度神经网络（Deep Neural Network，DNN）的方法，并对目前已有方法的缺点和不足 进行了简要总结与概括,主要包括:

1. 部分方法只使用到了局部特征，而忽略了局部特征与全局特征的结合。
2. 大部分方法未能考虑CTU中不同尺度的特征融合，，无法充分利用原始图像中 的多尺度信息。
3. 神经网络设计过于简单，不能够很好的提取CTU图像特征用于深度的划分。
4. 未能充分考虑量化参数QP对CU深度划分的影响。
5. 对于深度的划分采用串联二级分类方式，容易出现预测的错误累计效应。

综上所述，针对现有方法的缺陷，本论文提出了一种新颖的多尺度多输入卷积神经网络（Multi-scale Multi-input Convolutional Neural Network：MMCNN）。提出的MMCNN同时考虑了将多个QP值（本论文考虑了22、27、32和37四个不同的QP值）和多个不同尺度的CU图像块作为输入，通过对不同尺度CU之间特征信息的有效融合，进一步提升了网络的特征提取能力和表达能力，极大提升了HEVC帧内CTU深度划分的效率

除此之外，不同于现有方法的串联二分类策略，本论文还将多分类融入到CU的深度预测中，提出了一种互补的分类策略（Complementary Classification Strategy ：CCS）：首先，创建两个MMCNN来分别预测CTU中每个的CU的深度（三分类）和判断每个的CU是否需要划分（二分类）；然后，综合考虑三分类和二分类的结果，并采用投票机制来决定每个CU的最终深度结果。通过结合多分类和投票机制，CCS中的三分类和二分类很好的实现了分类结果的相互互补，很好地避免了现有方法中存在的错误累积效应，较大提升了CTU中CU深度预测的准确性。

## 论文组织结构

# 视频编码相关理论及神经网络简介



## 引言

随着互联网技术的飞速发展，人类的通讯信息已经从传统的文字模式逐渐转变为视频模式。如今，海量的视频信息已经像潮水般涌入我们的工作和生活的各个领域。尽管近年来网络带宽和传输能力的增长速度令人瞩目，但这仍然无法满足海量视频数据的传输和存储需求。因此，高效的视频信息压缩技术在过去、现在以及可预见的未来，都将是解决这一问题的重要技术手段。

本论文将详细探讨视频编码以及快速视频编码的相关理论，内容主要分为四个部分。首先，我们将对视频编码的基础知识进行深入的介绍，包括视频信号的统计特性以及可压缩性原理。其次，我们将详细阐述当前最流行的视频编码技术——HEVC编码框架，包括其编码结构、帧内预测模式以及帧内预测过程。接着，我们将介绍神经网络的基础知识。最后，我们将探讨视频压缩的评价标准，包括视频码率以及图像质量的客观评价指标和主观视觉效果。视频码率和图像质量评价是评估视频压缩效果的重要指标。本论文旨在为读者提供一个全面、深入的理解视频编码和视频压缩的理论和实践

## 视频编码基础

### 视频信号的统计特性

尽管视频图像种类繁多，内容千变万化，数据量大得惊人，然而，大量的统计实验表明，图像数据本身存在一些内在联系和统计规律。例如，图像的同一行相邻像素之间，相邻行像素之间，以及相邻帧的对应像素之间往往存在很强的相关性，即通常人们所说的存在大量的“冗余”信息。建立在信息论基础上的图像编码方法就是利用图像信号这种固有的统计特性，通过去除这些冗余信息来对图像数据进行压缩处理的。

### 视频图像的自相关性

从信息论的角度出发，可以通过对图像信息的一阶熵和高阶熵的分析，来确定图像信源的统计特性。但图像熵值的计算十分困难，它要预先知道图像的概率分布等统计参数，因而在实践中用得较多的还是图像的相关函数，因为它可以直接反映任意图像像素之间的关联程度，可以在统计平均的意义上来计算它们之间的相似程度。

#### 像素差值的自相关函数

1. 空域像素的差值分布

可以推想，由于一幅图像内相邻像素值之间的相关性很强，相邻像素值之差的统计分布应该有相当一部分集中在零附近。图2-1是对多幅实际图像的水平方向相邻像素差值信号进行统计得到的概率密度分布示意图。大量的图像数据统计表明，对于灰度范围为0～255的常见图像，差值信号绝对值的80%～90%落在0～20的范围内。这一统计得出的结论在预测法图像压缩中是非常重要的依据。

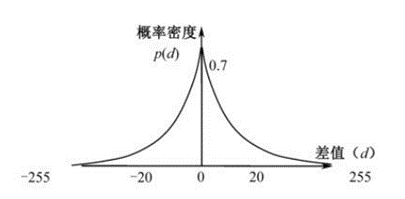


图2-1 图像差值信号的统计分布示意图

图像的差值的概率密度p（d）常用拉普拉斯（Laplace）分布来近似：

（2-1）

其中，为相邻像素值之差，为其均方差（标准差）值

1. 时域像素的差值分布

上述对一幅（帧）图像内部相邻像素或其差值进行的统计分析，通常称为空域（帧内）统计特性。对视频图像来说，相邻帧对应位置像素之间的时间间隔很小，只有几十毫秒，图像在这段时间内发生变化的可能性不大，变化的程度往往较小。以最简单的帧间差值为例，如图2-2所示，第k帧的帧间差定义为

(2-3)

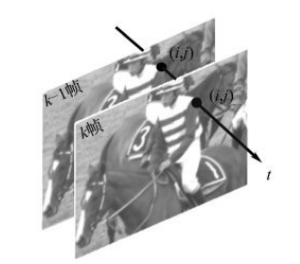


图2-2 帧间差位置示

其中，表示第帧处的一个像素，表示第帧和 处于同一位置的像素。一般认为，时域的差值信号的分布特性也和空域差值信号的分布特性类似，也服从拉普拉斯分布，仅具体的均值和方差有所差别。在视频图像中，除非景物有剧烈的活动，或整幅场景更换以外，相邻帧之间往往存在着很强的相关性。

正是由于视频图像的自相关性，其在空间和时间上存在大量的冗余信息。这些冗余信息在一定程度上增加了视频数据的大小，但并没有为视觉质量带来相应的提升。因此，通过有效地去除这些冗余信息，我们可以实现视频数据的压缩，从而在不显著降低视觉质量的前提下，减少视频数据的存储和传输需求。

### 混合编码

视频编码或视频压缩就是利用图像统计特性，去除视频信号中存在的大量冗余信息的过程。目前这种基于统计的视频压缩常见有两类方法，一类方法是建立在图像的差值信号分布集中基础上便于压缩处理的预测编码，如帧内编码、帧间编码，多视点视频中的视点间预测编码等。另一类方法是建立在正交变换可以将分散分布的图像数据，在变换域集中分布的基础上便于压缩处理的变换编码，如离散余弦变换、离散正弦变换、哈达马变换等。将这两种方法结合起来使用，即形成了混合编码，可消除图像数据中不同类型的冗余信息，取得比单独使用一种方法更高的压缩率。

#### 预测编码

预测编码（Prediction Coding）利用相邻像素的空间或时间相关性，用已传输的像素对当前正在编码的像素进行预测，然后对预测值与真实值的差——预测误差进行编码和传输。目前用得较多的是线性预测方法，即用已传像素的线性组合对正在编码的像素进行预测。

预测编码是图像压缩技术中研究得最早，且应用最广的一种方法，它的一个重要的特点是性能较好，算法简单，易于硬件实现。图2-3是一个空间预测编码的示意图，其中的处理单元主要包括线性预测器和量化器两部分。在这种编码方案中，编码输出的不是图像像素的样值，而是该样值与预测值之间的差值，即预测误差。据图像信号的统计特性的分析，可以得到一组最佳的预测系数，使得预测误差的分布大部分集中在0附近，经非均匀量化，然后传输量化后的预测误差信号。

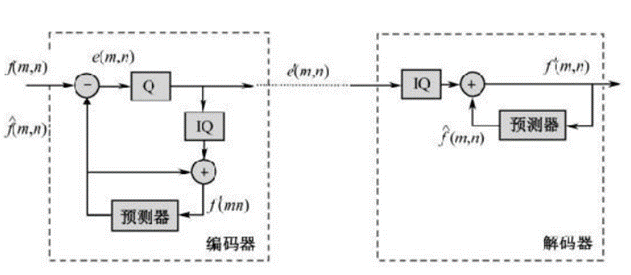


图2-3 预测编码原理框图

解码过程和预测编码相反，由解码器采用和编码器相同方法计算出的预测值 加上由编码器传输来的预测误差信号，即可得到重建的图像，它和输入视频的误差是由量化操作造成的。预测编码中的量化器可以根据人眼的视觉特性来设计，采用较少的量化分层，而量化噪声又不被人眼所觉察。这样，图像数据得到了压缩，而图像的主观质量并没有明显下降。如果在编解码器取消量化和反量化单元，则可形成一个无失真预测系统，解码恢复的图像和原图象完全一致，即有成立。但这种无失真系统的压缩效率要远低于有量化的有失真预测系统。

#### 变换编码

变换编码（Transform Coding）将空间域描述的图像，经过某种变换（如离散余弦变换、离散正弦变换、哈达玛变换等）形成变换域中的数据（系数），达到改变数据分布，减少有效数据量的目的

在变换编码中，正交变换是一种最常见的数据处理手段，它把统计上彼此密切相关的像素值矩阵通过线性正交变换，变成统计上彼此较为独立、甚至完全独立的变换系数矩阵。信息论的研究表明，正交变换不改变信源的熵值，变换前后图像的信息量并无损失，完全可以通过反变换得到原来的图像值。但经过正交变换后，数据的分布发生了很大的改变，变换系数在变换域坐标系分布趋于集中，如集中于少数的直流或低频分量的坐标点。数据的集中分布为数据压缩创造了条件。比如，有利于通过量化操作去除大部分零或接近零的系数，保留少量有效系数；有利于对量化后的系数采用更加有效的表示方式，如“之”（Zig-zag）字形扫描、变长编码等，从而获得对图像信息量的有效压缩。

视频编码中最常见的正交变换是离散余弦变换（Discrete CosineTransform,DCT）。在目前常用的正交变换中，DCT变换被认为是一种准最佳变换。DCT变换矩阵与图像内容无关，去相关性好，有快速算法（FDCT），实现方便；DCT的基函数是偶对称的数据序列，可减轻在图像的分块编码中块边界处的灰度值跳变和不连续现象。DCT这些优点，使得二维DCT变换在图像编码的应用中得到普遍的使用。在历年来颁布的一系列视频压缩编码的国际标准中，都把基于块的DCT作为其中的一个基本处理模块。

采用DCT变换的图像压缩编码基本框图如图2-4示。编码器根据DCT系数集中在低频区域，越是高频区域系数值越小的特点，利用人眼的视觉特性，通过设置不同的视觉阈值的量化电平，将许多能量较小的高频系数量化为0，可以增加变换系数中0的个数，同时保留能量较大的系数分量。对量化后的系数还可以再进行变长编码从而获得进一步的压缩。在变换编码的解码端，其解码过程正好和编码过程相反，可以获得失真很小的解码图像。这里失真仍然是由量化器引起，因为正反变换、变长编解码都是无失真处理。

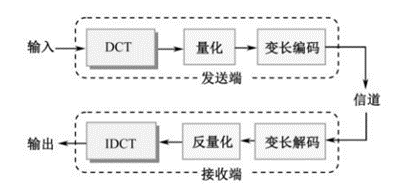


图2-4 变换编码的基本框图

## HEVC混合编码框架

从编码原理看，HEVC采用的是混合编码方式。它作为H.264/AVC的后继者，有着和H.264相似的结构，从某种程度上可以说HEVC是高级的H.264/AVC。HEVC 标准包括了 3 大模块，分别为预测编码、变换编码和熵编码等。其中预测编码和熵编码均为无损编码，变换编码中的量化是失真的主要原因，也直接决定了压缩率。HEVC的编码框架如图2-6所示

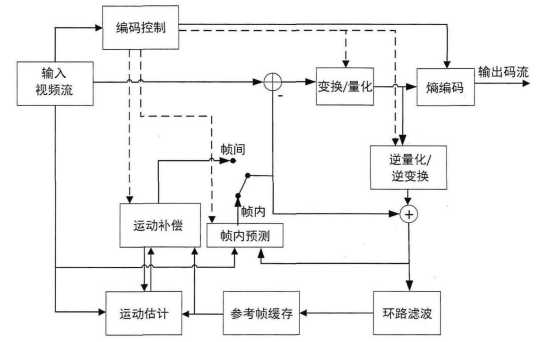


图2-5 HEVC编码框架

HEVC的编码框架是基于块的预测变换混合架构，当对输入视频序列进行编码时，首先对每帧图像划分成若干个预先设定大小的编码块。对输入信号编码的第一帧或者是随机切入点后的第一帧一定采用帧内预测，此过程利用空间相关性得到每一个编码块的预测值，不需要参考其他输入帧。视频序列中其他视频帧，则大部分采用帧间预测模式进行编码，帧间预测的两大关键点是运动矢量的确定和参考图像的选择。输入的视频序列信号经过帧内或帧间预测之后，用残差信号表示与原始图像存在的差异，该信号通过线性空间的变换之后再经过量化，传至熵编码器，最终形成编码码流传至解码器。与此同时，量化后的数据经过反量化、反变换、环路滤波等操作，形成重构图像，存在解码图像缓冲区中，参与后续图像的预测。

### HEVC编码结构

HEVC将一个视频序列分为相继的若干图像组（GOP），每一组由该序列中连续的多帧图像组成。帧是四叉树划分的基本单位，每一帧图像经过四叉树划分，形成覆盖全帧的多个同样尺寸的方形编码树块CTB（Coding Tree Block）。CTB还可以划分为更小的编码块CB（Coding Block）。CB是实施视频编码算法的基本单位，它还可以划分为预测块PB(Prediction Block)和变换块TB(Transform Block)。

同一位置的亮度 CTB 和两块色度 CTB，再加上相应的语法元素以及所包含的编码单元（Coding Unit， CU）形成一编码树单元（Coding Tree Unit，CTU）。

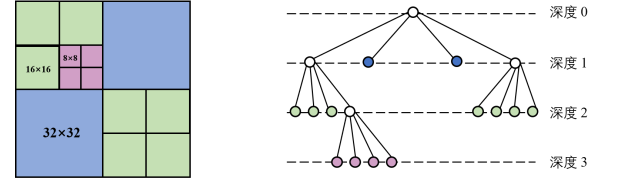
CTU可以按照四叉树结构分解为若干方形编码单元（CU），同一层次的CU必须是同一尺寸的4个方块，最多可有4层分解，即64×64,32×32,16×16和8×8。如果不分解，则这个CTU仅包含一个CU，此时亮度CTB的尺寸就是亮度CB的最大尺寸。

另外，由于图像中一定大小的CTU是按照光栅扫描的方式排列的，因此在图像的右方或下方边缘处，某些CTU覆盖的区域可能会部分地处于图像边界之外，这意味着CTU必须继续进行四叉树分裂，减小CB尺寸，直至最小允许的亮度CB尺寸，以使得所有的CB可完全安放到图像中

#### CTU的四叉树划分

CTU和CU之间的关系可表示为一棵四叉树，CTU为树的根节点。CTU还可以划分为更小的方块，相当于由根节点向下层分叉的4个分枝，每个分枝又可以继续分叉为下一层的4个分枝，直至不需再分的编码单元（CU），相当于编码树的最底层——树叶。CU的尺寸最大64×64到8×8，分别对应的深度是从0到3。此外，由编码器选择、并标注在比特流中的实际最大和最小的CU尺寸可施加更紧的限制。

CTU的划分是一个递归的过程，HEVC通过对所有可能的预测编码模式和划分深度，进行递归遍历的率失真优化，以找到最合适的 CTU 划分结果。具体来讲，HEVC会计算尺寸为64x64的CTU即深度为0时的率失真代价，然后计算此CTU划分为4个子CU即深度为1时的率失真代价，如果深度为0时的率失真代价比深度为1的率失真代价小，则停止划分，此时的CTU深度为0，CTU划分结束，否则，将每个深度为1的CU继续划分为4个子CU，即深度为2，计算此时的率失真代价。每个深度为1的CU与其划分出的4个子CU的率失真代价进行比较，如果深度为1的率失真代价小于深度为2 时的率失真代价，则该深度为1 的CU停止划分，否则继续将其进行划分，并计算率失真代价，直到划分的尺寸为8x8的CU则停止划分。如图2-6所示，CTU的划分过程



2-6 CTU划分示例

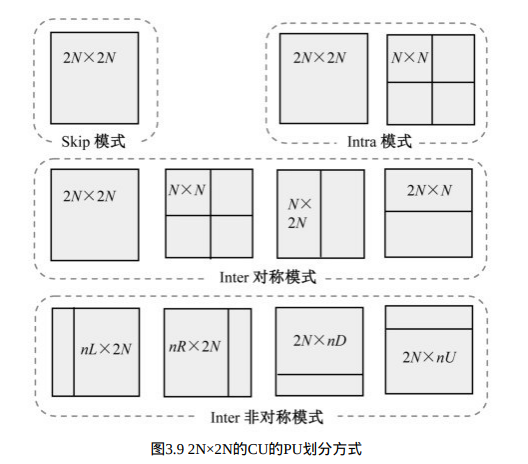
### HEVC帧内预测模式

HEVC的帧内预测技术和H.264/AVC类似，采用基于块的多方向帧内预测方式来消除图像的空间相关性，但比 H.264/AVC 预测方向更细、更灵活。HEVC 预测块的预测参考像素也和H.264/AVC类似，来自预测块左方的一列和上方的一行已编码的像素。HEVC为亮度预测块定义了33种不同的帧内预测方向，连同平面和直流模式，总共35种帧内预测模式。为了提高帧内预测的效率，HEVC对8×8或更大的预测块在预测前对参考像素进行了简单的平滑滤波预处理。

#### PU的预测模式

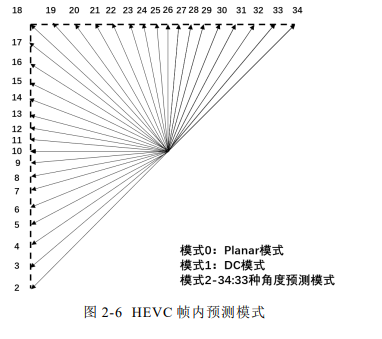
预测单元PU是HEVC进行预测运算的基本单元，只能定义在不再划分的最低层的CU中，包括帧内预测和帧间预测两类。CU决定了本单元包含的所有PU的预测方式和划分方式。

一个编码单元CU可以划分为一个或多个预测单元PU,CU到PU仅允许一层划分，最小的PU为4×4。划分可以是对称的，也可以是不对称的。如图3.9所示，一个2N×2N（N可以是4、8、16、32）的CU可划分为8种包含PU方式（图3.9中第2排和第3排）。帧间预测时可以在这8种方式中任意选择，帧间预测的跳过模式中只允许选择2N×2N这种方式，帧内预测只允许选择2N×2N或N×N方式。



2NX2N的CU的PU划分方式

为了匹配图像块的内容结构，提高预测精度，HEVC比H.264/AVC更加细分了预测角度，大大增加了预测方向。HEVC共有35种预测模式，其中33种为方向预测模式，或称角度预测模式。图4.6显示了亮度块预测33个方向的定义，连同直流和平面模式，共计35种预测模式，它们都有相应的编号



HEVC亮度块帧内预测的35种预测模式可分为如下的3大类：

（1）平面（Intra\_Planar）预测模式（模式0）平面（Planar）预测模式适合纹理比较平滑的区域，尤其是变化趋势比较一致的区域。Planar模式对当前块中每个像素使用不同的预测值，使用水平和垂直两个方向的线性插值的平均作为当前像素的预测值。（2）直流（Intra\_DC）预测模式（模式1）DC 模式对当前块的所有像素使用同一个预测值，即预测参考像素的平均值。这种模式适用于图像的平坦区域。

（3）角度（Intra\_Angular）预测模式（模式2～34）

HEVC的帧内角度预测和H.264/AVC类似，也是在空域的逐像素预测操作，但有显著的扩展，主要是PB尺寸的增加和可选预测方向数目的增加。使用Intra\_Angular模式时，每个帧内预测PU都有2个参考像素集：水平参考像素行——当前预测PU的上一行像素；垂直参考像素列——当前预测PU的左一列像素。预测PB的每个像素都可从那些在本方向预测前已经重建的参考像素集的样值进行角度预测。

### HEVC帧内编码算法流程

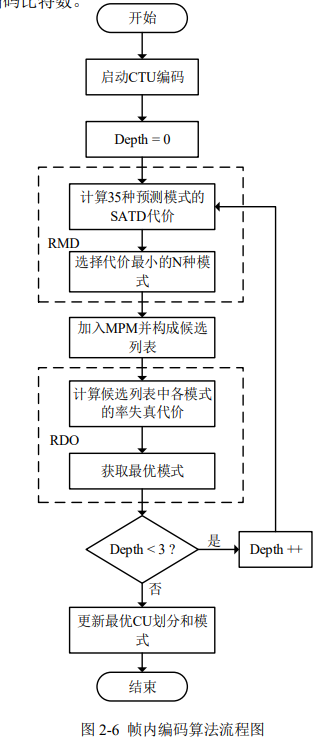
HEVC标准的帧内编码算法的具体流程如图2-6所示。在帧内预测编码的过程中，编码块CU的划分需要先从深度为0的编码块CU开始，逐渐往下划分到深度为3的编码块CU,对所有尺寸大小的编码块CU进行遍历。对于每一个编码单元CU,都需要对其35种帧内预测模式的计算过程进行遍历，先通过低复杂的粗略模式决策(Rough Mode Decision,RMD)算法，按照式(2-l)计算哈达玛代价(HadmardCost,HAD Cost)并进行排序，然后从中选择HAD Cost代价最小的N(N=3或8)个候选模式。



公式解释

然后将 N 个候选模式和最可能模式集（Most Probable Mode，MPM）中的预测模式构成候选列表，并分别对候选模式列表中的各个预测模式进行率失真优化（Rate Distortion Optimization，RDO）算法，按照式(2-2)计算率失真代价（RateDistortion Cost，RD\_Cost)并进行排序，选择 RD\_Cost 最小值对应的那个模式作为当前预测块的最优预测模式。





## 神经网络介绍 神经网络技术是 一种模拟人脑神经系统工作机制的计算模型，用于实现机器学习和人工智能。神经网络的基本组成单位是神经元，每个神经元接收一组输入，通过一个加权的方式将这些输入组合在一起，然后通过一个激活函数，生成一个输出。这个过程模拟了生物神经元接收和处理信息的方式。神经网络可以从海量数据中提取出有用的特征，并利用这些特征来训练模型、分类数据和预测未来事件。神经网络适合于建模非线性关系，通常用来进行模式识别，并对语音、视觉和控制系统中的对象或信号进行分类。神经网络是一种重要的机器学习技术，它可以解决预测复杂和需要深入理解的问题。

* 1. 卷积神经网络的基本结构

卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN)是一种深度学习的算法，它在图像和视频处理领域表现出了优异的性能。CNN的主要特点是能够自动、适应地学习局部区域的空间层次结构，这使得它在处理具有网格结构（如时间序列和图像）的数据时具有很高的效率。卷积神经网络（Convolutional Neural Networks，CNN）的基本结构主要由五个关键部分构成，包括输入层、卷积层、池化层、全连接层和输出层。接下来，我们将逐一详细解析这些组成部分的功能和作用。

输入层

输入层是卷积神经网络的第一层，负责接收原始的输入数据。对于图像来说，输入数据通常是一个三维的数组，包含图像的宽度、高度和颜色通道（RGB）。在输入层，我们通常会对原始数据进行一些预处理操作，其中一个常见的操作就是归一化（Normalization）。归一化的目的是将输入数据的各个特征调整到同一尺度，这样可以防止某些特征由于数值范围过大而对模型的训练产生过大的影响。对于图像数据，归一化通常是将像素值从原始的0-255范围，转换到0-1范围或者-1到1的范围。这样做可以使得模型的训练更加稳定，同时也可以提高模型的收敛速度。

除了归一化，输入层还可能进行其他的预处理操作，例如中心化（将数据减去均值）、白化（将数据的协方差矩阵转化为单位矩阵）等。这些操作都是为了改善数据的分布特性，使得模型能够更好地学习数据的特征。

卷积层

卷积层是卷积神经网络的核心组成部分，主要负责从输入数据中提取特征。这一层的名称来源于它的主要操作：卷积。卷积是一种数学运算，以下是计算公式



它通过一个称为卷积核（或滤波器）的窗口在输入数据上滑动，每次滑动都会对窗口覆盖的部分进行特定的运算，然后将结果存储在新的特征图中。这个过程会在整个输入数据上重复进行，直到生成完整的特征图。在卷积神经网络中，卷积层通常会有多个卷积核，每个卷积核都能够提取出一种特定的特征，例如边缘、角点、纹理等。这些卷积核是在模型训练过程中自动学习得到的。

卷积层的一个重要特性是参数共享，也就是说，同一个卷积核在整个输入数据上滑动时，都使用相同的权重。这大大减少了模型的参数数量，同时也使得模型具有平移不变性，即无论特征在图像的哪个位置，卷积核都能够检测到。

池化层

在卷积层提取出原始图像的特征后，我们通常会使用池化层来进行下一步的处理。池化层的主要作用是降低数据的维度，也就是进行降采样，以减少网络的计算复杂度。

池化层的工作方式是，它在输入数据上使用一个固定大小的滑动窗口，然后在这个窗口内进行特定的运算，如取最大值（最大池化）或者计算平均值（平均池化），然后将这个结果作为该窗口在输出特征图中的对应位置的值。这个过程在整个输入数据上重复进行，从而得到一个尺寸更小的输出特征图。

此外，池化层还有一个重要的特性，那就是它可以提高模型的空间不变性。也就是说，即使图像发生了微小的平移或旋转，池化层的输出往往不会有太大的变化，这使得模型能够更好地识别出图像中的对象，而不会受到这些微小变化的影响。

激活层

激活层（Activation Layer）在神经网络中起着非常重要的作用。它的主要功能是引入非线性因素，使得神经网络可以拟合复杂的非线性关系。

在神经网络中，每个节点的输入和权重的线性组合后，通常会通过一个激活函数进行非线性变换，然后输出到下一层。这个激活函数就是激活层的主要组成部分。

常用的激活函数有很多种，如Sigmoid函数、Tanh函数、ReLU函数（Rectified Linear Unit，修正线性单元）、Leaky ReLU函数、PReLU函数（Parametric ReLU）、ELU函数（Exponential Linear Unit）等。

其中，ReLU函数是最常用的激活函数之一，它的形式非常简单：f(x) = max(0, x)。也就是说，如果输入值大于0，那么就直接输出这个值；如果输入值小于或等于0，那么就输出0。尽管ReLU函数非常简单，但它在实践中的效果非常好，尤其是在深度神经网络中。

全连接层

全连接层（Fully Connected Layer，也常被称为FC层或者Dense层）是神经网络中的一种基本层，它的主要作用是进行高级特征的学习和最终的分类或回归任务。全连接层的工作原理是：它将前一层的所有输出节点与当前层的每一个节点都连接起来，也就是说，前一层的每个节点都会对当前层的每个节点产生影响。这就是“全连接”这个名字的由来。在卷积神经网络中，全连接层通常位于网络的最后几层。在前面的卷积层和池化层中，网络从原始图像中提取出了大量的局部特征，然后在全连接层中，这些局部特征会被整合起来，学习出更高级的、全局的特征。

输出层

输出层是神经网络的最后一层，它的主要作用是输出网络的最终预测结果。

输出层的节点数通常取决于具体的任务。例如，对于二分类问题，输出层通常只有一个节点，输出值表示正类的概率；对于多分类问题，输出层的节点数通常等于类别数，每个节点的输出值表示对应类别的概率；对于回归问题，输出层通常只有一个节点，输出值就是预测的连续值。

输出层的激活函数也取决于具体的任务。对于二分类问题，通常使用Sigmoid函数将输出值转化为概率；对于多分类问题，通常使用Softmax函数将输出值转化为概率分布；对于回归问题，通常不使用激活函数，或者使用恒等函数。

损失函数

损失函数使用主要是在模型的训练阶段，每个批次的训练数据送入模型后，通过前向传播输出预测值，然后损失函数会计算出预测值和真实值之间的差异值，也就是损失值。得到损失值之后，模型通过反向传播去更新各个参数，来降低真实值与预测值之间的损失，使得模型生成的预测值往真实值方向靠拢，从而达到学习的目的。

反向传播

神经网络的反向传播是用于训练神经网络的主要方法。反向传播的目标是通过调整神经网络中的权重和偏置，以最小化网络输出和实际目标之间的误差。以下是反向传播过程的详细步骤：

首先，我们需要进行正向传播，得到神经网络的输出。这个输出会与实际的目标值进行比较，以计算出误差。误差是一个标量值，它反映了神经网络的输出与目标值之间的差距。误差越大，说明神经网络的性能越差。

然后，我们需要计算误差关于每个权重和偏置的梯度。梯度是一个向量，它指向误差函数在当前点增长最快的方向。通过计算梯度，我们可以知道应该如何调整权重和偏置，以使误差最小化。在神经网络中，我们通常使用链式法则来计算梯度。

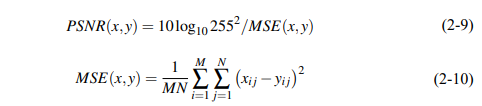
接下来，我们需要更新神经网络中的权重和偏置。更新的方式是将当前的权重和偏置减去学习率乘以对应的梯度。学习率是一个超参数，它决定了神经网络学习的速度。学习率过大，可能会导致神经网络在学习过程中震荡不定；学习率过小，可能会导致神经网络学习过程过慢。

然后，我们需要重复上述过程，直到神经网络的性能达到满意的程度。在每一次迭代中，我们都会使用一批新的数据来进行正向传播和反向传播，这就是所谓的随机梯度下降法。

## 评价指标

为了客观评价本论文所提出方法的有效性和先进性，在本文的所有实验中通过使用BD-BR（Bjφntegaard delta bit-rate）表示在同样的客观质量下，两种方法的码率节省情况，BD-BR的值越小越好）和编码时间节省率∆T以及PSNR，三种客观指标来定量评估本论文提出的方法的性能。

其中PSNR是最常被使用和应用的客观测量方法，用于衡量图像质量，它是按照重建图像与原始图像所有像素之间的均方误差（Mean Square Error，MSE）来定义的。对于尺寸均为 M ×N 的灰度级（8 位）发送端原始图像 x 和接收端恢复图像 y，x 和 y 之间的 PSNR 的计算公式一般如下：



BD-BR（Bjφntegaard delta bit-rate）是一种用于衡量视频编码效率的指标，由Kjell Bjøntegaard提出。它通过比较两种不同编码技术的比特率-失真曲线，计算出在相同失真水平下，一种编码技术相对于另一种编码技术可以节省的比特率百分比。如果BD-BR值为负，说明新的编码技术比旧的编码技术更有效率；如果BD-BR值为正，说明新的编码技术比旧的编码技术效率更低。

∆T表示加速算法相对于hevc的时间节省率

## 本章小结

# 基于多尺度网络的CU快速划分算法



## 本章引言

深度学习神经网络，特别是卷积神经网络（Convolutional Neural Network：CNN）在过去的几年中已经在优化HEVC帧内模式的CTU深度划分中发挥了重要的作用。这些基于CNN的深度预测方法通过自主学习CTU中CU的结构，代替了传统的计算RDcost方式，从而大大提高了CTU的划分效率，实现了编码的加速。然而，现有的基于CNN的CTU深度划分方法大多只考虑了单一尺度的输入，忽略了不同尺度CU之间特征的相关性。这是一个重要的问题，因为小尺度CU的结构特征会对包含了该小尺度CU的大尺度CU的深度划分产生极大的影响。因此，现有方法的性能仍有较大提升空间。

鉴于此，本章提出了一种多尺度多输入网络的方法，旨在解决上述问题。首先，我们设计了多尺度的输入，以便网络可以提取到更多的信息，从而克服了现有基于CNN的CTU深度划分方法忽略不同尺度CU之间的特征关联的问题。其次，我们注意到现有的基于深度神经网络的方法，都没有足够的关注QP对CU划分的影响，仅仅是为不同的QP训练出不同的网络权重，这种方法不具有泛化性。因此，我们将QP作为网络的学习参数输入到网络进行训练，以提高网络的泛化能力。最后，我们构建了三级串联的而分类网络，用于分别预测64x64、32x32和16x16尺寸CU是否需要划分。

总的来说，本章的主要贡献在于提出了一种多尺度多输入网络的方法，这种方法能够更好地处理不同尺度CU之间的特征关联，提高网络的泛化能力，并通过构建三级串联的而分类网络，有效地预测不同尺度CU是否需要划分。我们相信，这种方法将为优化HEVC帧内模式的CTU深度划分提供新的思路。。

## CU划分分析

在对HEVC编码软件对编码单元的划分实例进行大量分析后，我们发现帧内编码树单元的划分结构与视频图像的纹理复杂度和QP有着密切的关系。具体来说，对于纹理变化不大，即平坦区域的图像，HEVC通常会选择使用较大的CU进行编码，也就是对该CTU进行较少的划分，甚至不进行划分。这是因为在对背景单一，纹理简单的区域进行CU划分时，HEVC更倾向于选择深度小的编码单元对图像帧进行编码，以此提高编码效率。

相反，对于纹理复杂，细节丰富的区域，HEVC会选择使用较小的CU进行编码，也就是对该区域进行深层次的划分，通常会划分两次或三次。这是因为在这种情况下，使用较小的CU进行编码可以更好地保留图像的细节信息，从而提高编码的质量。

为了更直观地说明HEVC的划分规律，我们选取了一幅最具代表性的图片进行展示。从图中可以清楚地看出，对于背景简单，平滑且没有纹理细节的区域，HEVC会使用较大的块进行划分，通常划分的深度在0 或1，而对于纹理丰富，细节变化大的区域，HEVC则会对其进行更深层次的划分，通常划分的深度在2或3层。



CU划分实现

。

此外，我们还对不同深度的CU在视频编码过程中的占比进行了定量分析。我们使用了一种算法对CU的深度划分进行统计分析。通过对HEVC编码软件HM代码的分析，我们发现，HM在编码过程中会对CTU进行递归的划分，并存储了划分的深度数据。对于尺寸为64x64的CTU来说，其中的每4x4个像素块会被记录一个深度信息，所以需要一个16x16的矩阵进行存储。但是，由于CU被允许划分的最小尺寸为8x8，即8x8的块覆盖下的4x4的像素块所处的深度应该是一样的，所以只需要一个8x8的矩阵就可以存储整个CTU的深度划分信息。

|  |
| --- |
| 算法**1 投票算法** |
| **输入：**视频序列和编码配置信息  **输出：深度为 0,1,2,3的CU占比**   1. for I <totalNumberPart do 2. depth = getDepth(i) 3. if depth = j ,j[0,3] do 4. count[j]++ 5. end if 6. percent[j] = count[j]/totalNumberPar |

算法 1 在 H.265/HEVC 官方编解码软件 HM16.25 上实现，采用帧内编码，量化参数设置为 22、27、32、37， 帧内编码配置文件为encoder intra main.cfg，

所用的测试序列是通用测试条件规定的序列 Traffic､BQTerrace､PartyScene 和 FourPeople。

由以上图表数据分析可知，不同分辨率下的深度占比并没有明显的规律，但是在同一视频序列的不同QP中却有着明显的规律。对于同一视频序列，随着QP的递增深度为0 和1 的占比都在增加。可以看出除了partScenc序列，所有的视频序列对于深度0的占比traffic 1.8%升高至9.3，bQtrrrac 由8.6升高致15.7，fourPeople 由2.1 升到11.2 ，对于深度0的占比traffic 1.8%升高至9.3，bQtrrrac 由8.6升高致15.7，fourPeople 由2.1 升到11.2

通过对HEVC编码单元划分实例的深入分析，我们发现了一个显著的趋势：随着量化参数（QP）的增大，编码单元（CU）划分深度为0或1的占比会以不同程度的方式增加。同时，深度为2的CU划分占比呈现出上升或持平的状态，而深度为3的CU划分占比则持续下降。这一发现表明，随着QP的增大，HEVC编码器更倾向于将视频内容划分为较大的块，即CU划分的深度更浅。相反，对于较小的QP值，HEVC编码器会更精细地进行CU划分，倾向于划分出较小的块。

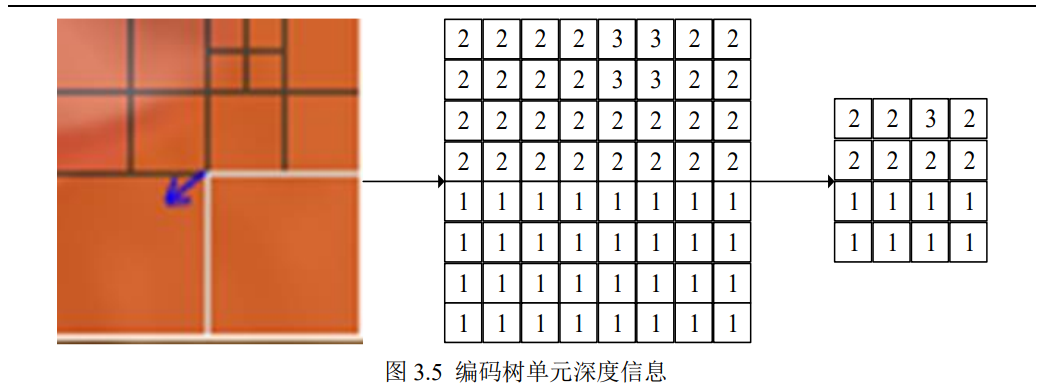
基于这一观察，我们提出了一种创新的方法，将QP作为神经网络的学习参数。这种方法的优点是双重的。首先，它增加了神经网络的学习性能。通过将QP纳入学习过程，神经网络能够更好地理解和模拟HEVC编码器的行为，从而提高其预测精度。其次，这种方法也增强了神经网络的泛化能力。由于QP是影响视频编码质量和效率的关键参数，因此，训练出的神经网络模型能够适应不同QP设置下的视频编码任务，具有更强的泛化能力。

## 数据集制作

在本项研究中，我们构建了一个专门为神经网络训练而设计的帧内CTU划分数据集。该数据集的数据来源于CPIH[23]，包含了4928×3264、2560×1600、1920×1080以及1280×720四种不同分辨率的视频序列。这些丰富的数据源为我们的神经网络训练提供了充足的数据保障。为了获取编码后的CTU划分结果我们将这些视频序列使用HEVC标准编码器HM16.25在四中不同QP{22,27,32,37}下进行帧内模式的编码，然后获得编码后的CTU划分结果。需要注意的是，CTU的亮度分量包含了大量的纹理信息，这些信息可以很好地反映当前CTU的图像分布，而CU的划分也是利用这些纹理信息进行的。

在数据集的制作过程中，我们保存了CTU的亮度分量、编码时使用的QP值以及编码后的CTU结果。具体的制作方法如下：

首先，我们将所有的视频序列在4中qp中分别使用帧内模式进行编码。在编码结束后，每个编码的CTU最终将会得到一个8x8的矩阵，用于存储划分的深度信息。



如图所示深度划分结果，CTU中每个16x16尺寸的CU在8x8矩阵中对应一个深度信息。对于64x64的CU，如果矩阵中的深度值都不为0，那么就代表该CU进行了划分，我们将其对应的标签设为1，反之如果全为0就代表该CU未划分将其标签设置为0.对于32 x32尺度的CU,如果其包含的深度信息全为1，则代表该32x32的CU不划分，如果包含的深度信息为2或3则代表该32x32的CU进行了划分。对于16x16尺度的CU,如果其对应的深度信息为2，则代表该16x16的CU不划分，如果包含的深度信息为3则代表该16x16的CU进行了划分。

最后我们将三种不同尺寸的CU和其对应的编码时QP和二分类标签，分为三个数据集，以方便不同层级预测网络的训练。

## 三级预测网络构建

在高效视频编码（HEVC）中，编码树单元（CTU）的划分过程是从64x64的尺寸递归划分到8x8的编码单元（CU）。这个过程需要经历复杂的率失真优化（RDO）计算，这也是HEVC编码复杂度高的主要原因。为了解决这个问题，我们受到HEVC的划分流程的启发，将CU的划分问题转化为一个二分类问题，并设计了一个三级预测分类网络。这个网络可以直接预测CU的划分情况，从而替代了标准编码器的RDO过程。这种方法在保持编码性能的前提下，有效地降低了编码复杂度。

具体来说，我们将CTU的划分过程转化为三级二分类问题。首先，对于尺寸为64x64的CU，我们需要判断其是否需要划分为4个32x32的CU。然后，对于每个32x32的CU，我们需要判断其是否需要划分为4个16x16的CU。最后，对于每个16x16的CU，我们需要判断其是否需要划分为4个8x8的CU。在这三次划分过程中，我们设计了三个预测网络，分别用于预测每种尺寸CU的划分情况。我们使用net1来预测64x64的CU是否需要划分，使用net2来预测32x32尺寸的CU是否需要划分，使用net3来预测16x16尺寸的CU是否需要划分。

在具体的预测过程中，首先将64x64的CU块输入到net1中，获得其预测的划分结果。如果预测结果为0，那么代表64x64的CU不需要划分，将这个预测结果保存下来。如果64x64的CU预测需要划分，那么将64x64中的4个32x32的CU依次输入到net2中，进行预测32x32的CU是否需要划分，并保存预测结果。如果有32x32的CU的预测结果为1 即需要划分，那么将32x32的CU下的4个16x16尺寸的CU输入到net3中，进行预测其是否需要划分，并保存预测结果。

在得到三级预测结果后，将这些结果转化为CTU深度信息的8x8矩阵，以供编码器进行下一步处理。具体的转化方法如下：如果在net1中的预测结果为0，即64x64的CU不需要划分，那么整个深度信息矩阵的值全为深度0。如果net1的预测结果为1，即64x64的CU需要划分，那么先将深度信息矩阵的值全设为深度1。然后，我们判断4个32x32的CU在net2的预测结果，如果有32x32的CU预测需要划分，那么将对应位置的值设置为深度2。最后，将预测需要划分的32x32的CU中的4个16x16的CU输入到net3中，预测其是否需要划分。如果预测结果是1，即16x16的CU需要划分，那么将8x8的深度信息矩阵对应位置设置为深度3。至此，整个CTU的深度信息预测就完成了。



CTU分区预测流程图



64x64预测

****

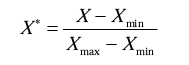
32x32预测



16x16预测

本章提出的基于多尺度网络的CU快速划分算法主要包括一级预测结构、二级预测结构以及三级预测结构，其中，一级预测结构用于预测64×64CU是否需要划分，二级预测结构用于预测32×32CU是否需要划分；三级预测结构用于预测16×16CU是否需要划分，这三个神经网络都包含输入层、卷积层、池化层和全连接层、以及输出层。

在输入层对原始的视频图像数据进行去均值、归一化等预处理操作[38]。这是因为归一化可以帮助改善模型的训练速度和性能。归一化处理可以将输入数据转换到一个公共的尺度上，使得不同特征之间的数值范围相近，这样可以防止因为数值范围过大或过小而导致的梯度消失或梯度爆炸问题。同时，它们也可以帮助减少模型对于图像亮度和对比度的敏感度，使得模型更加关注图像的内容。神经网络的中心思想是拟合地编码单元的划分与视频图像内容的关系。本章采用视频图像像素表示视频内容，输入的编码树单元信息是从YUV格式的原始图像或序列中提取的，在神经网络中只使用Y通道，因为该通道包含了最多的视觉信息[39]。神经网络网络对输入的亮度信息和量化参数进行归一化操作，计算公式如下



其中，X为像素原始值，Xmin为像素最小值，Xmax为像素最大值，X\*为经过归一化处理后 的值，

*卷积层*

卷积层的一个重要特性是参数共享，即在整个输入数据上，我们都使用同一个卷积核。这大大减少了模型的参数数量，同时也使得模型具有平移不变性。再本章的3个网络中 我们分别使用了5x5和3x3的卷积核对特征图进行卷积操作



其中a aa代表输入图片，b bb代表输出特征图，w ww是卷积核参数，它们都是二维数组

经过卷积运算后的输出与输入图片的尺寸不同，3个超参数控制着输出的尺寸：深 度，步长和零填充[40]。输出数据的深度与所用的卷积核数目相匹配，不同的卷积核在输 入数据中所学到的特征是不一样的。滑动图片进行卷积运算，当步长为n时，每次移动 n个像素。滑动至图片边缘时，若图片剩余区域与卷积核大小不匹配，需要用0在边缘 处进行填充，称为零填充。卷积后输出尺寸的计算公式为：



其中，H是输出尺寸，N是输入图片的尺寸，F是卷积核的大小，P为填充值，S是卷积

步长

*池化层*

在神经网络中，池化层起着至关重要的作用，它对从特征提取模块得到的特征图进行下采样操作，以达到降低信息冗余、提升模型的尺度不变性和旋转不变性，以及防止过拟合的目的。池化层的工作原理是通过在特征图上滑动一个预定义的池化窗口，然后对窗口覆盖的特征值进行某种池化操作，从而得到一个新的下采样特征图。这个过程可以被视为一种特征压缩和抽象的过程，它可以有效地减少数据的维度，从而降低模型的复杂度和计算量。常见的池化操作有最大池化（Max Pooling）和平均池化（Average Pooling）。最大池化是取池化窗口覆盖的特征值中的最大值作为输出值，而平均池化则是取平均值。在本章中全部采用的是最大池化，其公式可表示为Y(i, j) = max(X[ai : a(i+1), bj : b(j+1)])

*全连接层*

*在经过了卷积操作和最大池化操作，将转化为向量的数据输入到全连接层，在全连接*层中的每一个神经元都与前一层中的所有神经元相连。这种全连接的特性使得全连接层能够获取前一层所有神经元的信息，从而实现特征的全局整合。学上，全连接层的计算可以表示为一个矩阵乘法和一个加法。假设前一层的输出为向量x，全连接层的权重矩阵为W，偏置向量为b，那么全连接层的输出y可以表示为：

Y= Wx + b

其中，Wx表示矩阵W与向量x的乘积，"+"表示向量的加法。

*输出层*

在本章的方法中，将CTU分区问题转化为三级二分类问题，所以在网络的输出层采用softmax函数进行二分类， 输出结果表示输入CU划分和不划分的概率。

其他层： 在网络训练阶段， 为防止过拟合， 在第每个卷积层之后采用了dropout方法 神经元以50%概率失活。 网络中所有卷积层和隐藏的全连接层都以ReLu作为激活函数.

一级预测结构中包括单尺度特征提取模块、深层特征提取模块和全连接模块以及CU分区预测输出层，其中，单尺度特征提取层对应输入64×64CU的亮度分量 ；

二级预测结构中包括双尺度特征提取层、深层特征提取层和全连接层以及CU分区预测输出层，其中，双尺度特征包括第一层和第二层，第一层对应输入64×64CU，第二层对应输入32×32CU；

三级预测结构中包括多尺度特征提取层、深层特征提取层和全连接层以及CU分区预测输出层，其中，多尺度特征包括第一层、第二层以及第三层，第一层对应输入64×64CU，第二层对应输入32×32CU；第三层对应输入16×16CU。

接下来以预测CU的网络net3为例 介绍网络的主体架构。

本章提出的用于预测CU是否需要划分的三级预测结构如图1所示，首先在输入端用了多尺度输入的策略，多尺度特征提取层设置为三层，共有三个尺度对应三层特征提取层，每一层分别对应不同尺寸的CU：第一层对应的CU，第二层对应的CU，第三层对应的CU；将的CU输入到第一层，同时将该CU中所有的CU和所有的CU分别对应到第二层和第三层的输入。这是因为对于16x16CU，由于其自身的信息较少，网络很难学习到有用的信息对CU进行划分，又因其作为32x32CU的组成部分，其划分受到父CU的影响，因此将32x32的CU一并输入到网络，同理，64x64的CU作为包含整体信息的块也一并输入到网络中，经过多尺度特征提取模块，将64x64CU 变换为32x32分辨率的特征图，32x32CU 变换为16x16分辨率的特征图16x16CU 变换为8x8分辨率的特征图具体过程如下：

其中代表最大池化函数，代表卷积函数，代表输入的CU图像数据其中 代表CU的尺寸，分别为 64、32、16，代表特征图的分辨率分别为32、16、8.

表1

|  |  |  |
| --- | --- | --- |
| Layer | Filter Size | Number of Filters |
| Conv-64-1 | 5x5 | 16 |
| Conv-64-2 | 3x3 | 24 |
| Conv-64-3 | 3x3 | 32 |
| Conv-32-1 | 3x3 | 16 |
| Conv-m | 3x3 | 56 |
| Conv-m1 | 3x3 | 128 |
| Conv-m2 | 3x3 | 64 |
| F1 | 1x1 | 64 |
| F2 | 1x1 | 32 |
| F3 | 1x1 | 16 |
| F4 | 1x1 | 8 |
| Out | 1x1 | 2 |

如表1示出的用于预测CU是否需要划分的三级预测结构的网络参数配置表，对于三级预测结构而言，包括如下操作：

在多尺度特征提取中，我们对16x16和32x32的CU采用了步长为1 尺寸为3x3的卷积核进行卷积处理，并在该卷积层之后执行1次2 的最大池化操作在保留显著特征的同时实现降维操作。对于64x64的尺度，使用了5x5的卷积核进行卷积操。

在之后的深度特征提取层中，将在特征提取模块中对64x64操作后的特征图先采用2个3×3的卷积层来进一步提取细节特征，其中，第1个3×3卷积层的步长为1，第2个3×3卷积层的步长为2，将特征图的分辨率降为16×16,将此次的特征图与32x32CU进行特性提取后的同分辨率特征图进行通道上的拼接，组成新的特性图

接着将拼接后的特征图进行一次卷积和池化操作提取其特征后降维到8x8的分辨率并与多尺度提取模块中的8x8分辨率的特征图进行拼接，将再次拼接后的特征图经过2个步长为1的3×3卷积层和2个2×2的最大池化操作，最终得到一个维度为2×2×64的特征图，将其转换为向量后输入到全连接层进行处理

在全连接层使用了4个隐藏层，在前三个隐藏层中利用当前CU编码时的QP信息作为1个神经元和隐藏层连接起来， 为当前CU划分提供先验信息。

对于预测32×32的CU是否需要划分的网络结构如图2所示，同样是包括多尺度特征提取层、深层特征提取层和CU分区预测输出与图1 网络不同的是在多尺度上面，我们取消了16×16尺度的输入，保留了64×64、32×32的两个尺度。因为我们在图2网络中的是用于预测32×32尺寸的CU是否需要划分。所以网络只需要学习32×32尺度的CU特性就可以了

在预测64×64的CU是否需要划分时，我们设计了如图3所示的网络结构，与前两个网络不同的是，我们仅输入了64×64的尺度提供给网络进行学习

在这三个网络中我们都采用交叉熵作为其损失函数，如公式(1)所示：

在公式(1)中，N表示不同尺度下CU的数量，对于一个CTU来说，只需要一次判断的CU是否需要划分，所以此时N=1，的CU又包含4个的CU，所以在图2的网络中N=4，每个 的CU又包含4个 的CU所以在图1中N=16。 和分别表示每个CU真实的类别和网络输出的类别。i表示类别数，由于我们需要预测是每个CU是否需要划分，是一个二分类问题，所以 i∈[0,1]。

## 实**验结果及分**析

### 实验设置

我们把在CPIH数据集上训练好的模型直接用到JCT-VC（Joint Collaborative Team on Video Coding）的18个标准视频序列中进行编码复杂度的评估。在标签的制作方面，我们采用HEVC参考软件HM16.25（HEVC Test Model），其中所有的帧内配置模式都采用默认配置文件encoder\_intra\_main.cfg，并考虑了四种不同的QP值：22,27,32和37，来对CPIH和JCT-VC进行帧内编码，将得到的CTU深度划分结果作为每个CU的真实深度值（Ground Truth），即标签。本发明采用分类准确率（即：在网络输出的所有CU深度值中，判断正确的占比，其值越高越好）、BD-BR（Bjφntegaard delta bit-rate）表示在同样的客观质量下，两种方法的码率节省情况，BD-BR的值越小越好）和编码时间节省率∆T，三种指标来定量评估本发明提出的MCCN的性能，其中ΔT的计算如公式（2）所示：

ΔT=(T\_HM-T\_test)/T\_HM (2)

其中，T\_HM是HM16.25所需的时间，T\_test是本发明方法所需的时间。因此，ΔT越大越好，表示节省的编码时间越多，即编码复杂度越低。

在训练阶段，针对每个网络采用相同的实验设置：训练共执行200次epoch，批次大小设置为1024，采用随机梯度下降优化器Stochastic Gradient Descent（SGD）[25]，动量设置为0.9，初始化学习率为0.01，并在每个epoch之后以固定步长0.5×10^(-4)进行下降。本发明的所有实验都是在 Intel(R) Core(TM) i7-7820X CPU @ 3.60GHz和 NVIDIA GeForce RTX 2080 Ti的windows10 64位操作系统的计算机，以及Pytorch深度学习框架上进行训练和测试的。

### 实验结果

本章所提出算法的实验结果在表3-2中展示其中Class代表序列类别，Sequence是视频序列名称。表中给出了相对于HEVC原始编码器HM的性能变化，包括BDBR、BDPSNR以及编码时间节省通过时间节省率来比较两种算法的编码复杂度。从表中可以看出，所提算法对于不同分辨率大小的视频序列都表现了较好的性能，都很大程度上减少了编码时间复杂度，尤其是序列Kimono、Johnny、KristenAndSara平均编码时间节省都达到了79%以上，分别为79.73%、79.10%、79.05%，而BDBR、 BDPSNR的性能损失几乎可以忽略。所有序列在四种QP{22,27,32,37}下的平均时间节省分别为66.8%、68.24%、68.16%、68.90%，可以看出所有的视频都极大的减少了原始编码器中存在的冗余计算。同时可以发现随着QP参数值的增加，编码时间节省也随之增多，说明当QP值较大时，所提算法能够相对来说提供更多的时间节省。另外，所有测试序列在四种不同QP下的整体编码时间节省平均值也高达68.90%，最高可以达到82.51%的时间节省，而BDBR仅增加了2.27%，BDPSNR仅降低了0.16dB,以此极大证明了以本章所提出的三级预测结构的快速CU算法替代原始编码的RDO过程的有效性和可行性。

M-CNNS 算法实验结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sequence | BDBR(%) | BDPSNR(dB) | TS(%) | | | | |
| Qp=22 | Qp=27 | Qp=32 | Qp=37 | Average |
| A | Traffic | 3.10 | -0.17 | -69.84 | -69.78 | -70.10 | -70.45 | -70.04 |
| PeopleOnStreet | 2.40 | -0.14 | -69.79 | -69.52 | -69.72 | -69.88 | -69.73 |
| B | Kimono | 11.92 | -0.40 | -77.27 | -79.51 | -82.51 | -79.61 | -79.73 |
| ParkScene | 2.38 | -0.10 | -70.52 | -70.55 | -70.89 | -70.95 | -70.73 |
| Cactus | 2.63 | -0.10 | -66.29 | -73.35 | -73.72 | -74.14 | -71.87 |
| BasketballDrive | 6.99 | -0.19 | -74.30 | -69.14 | -57.69 | -62.37 | -65.87 |
| BQTerrace | 1.78 | -0.10 | -71.00 | -70.33 | -67.82 | -62.09 | -67.81 |
| C | BasketballDrill | 4.20 | -0.20 | -56.34 | -66.31 | -56.81 | -68.32 | -61.95 |
| BQMall | 2.21 | -0.13 | -66.71 | -66.51 | -66.06 | -67.92 | -66.80 |
| PartyScene | 1.45 | -0.11 | -59.45 | -59.31 | -59.52 | -59.79 | -59.52 |
| RaceHorses | 2.41 | -0.16 | -68.18 | -66.92 | -67.42 | -68.13 | -67.67 |
| D | BasketballPass | 2.43 | -0.14 | -50.72 | -51.82 | -66.57 | -68.06 | -59.30 |
| BQSquare | 1.61 | -0.14 | -60.05 | -59.73 | -59.57 | -60.27 | -59.90 |
| BlowingBubbles | 1.74 | -0.11 | -60.49 | -60.05 | -59.96 | -60.44 | -60.23 |
| RaceHorses | 1.96 | -0.13 | -65.51 | -64.33 | -64.21 | -63.75 | -64.45 |
| E | FourPeople | 3.14 | -0.18 | -64.33 | -72.74 | -72.54 | -73.49 | -70.77 |
| Johnny | 3.50 | -0.15 | -73.26 | -79.75 | -82.43 | -80.97 | -79.10 |
| KristenAndSara | 2.93 | -0.15 | -78.43 | -78.73 | -79.39 | -79.64 | -79.05 |
| Average | | 3.27 | -0.16 | -66.80 | -68.24 | -68.16 | -68.90 | -68.03 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sequence | BDBR(%) | BDPSNR(dB) | TS(%) | | | | |
| Qp=22 | Qp=27 | Qp=32 | Qp=37 | Average |
| A | Traffic | 3.10 | -0.17 | -69.84 | -69.78 | -70.10 | -70.45 | -70.04 |
| PeopleOnStreet | 2.40 | -0.14 | -69.79 | -69.52 | -69.72 | -69.88 | -69.73 |
| B | Kimono | 11.92 | -0.40 | -77.27 | -79.51 | -82.51 | -79.61 | -79.73 |
| ParkScene | 2.38 | -0.10 | -70.52 | -70.55 | -70.89 | -70.95 | -70.73 |
| Cactus | 2.63 | -0.10 | -66.29 | -73.35 | -73.72 | -74.14 | -71.87 |
| BasketballDrive | 6.99 | -0.19 | -74.30 | -69.14 | -57.69 | -62.37 | -65.87 |
| BQTerrace | 1.78 | -0.10 | -71.00 | -70.33 | -67.82 | -62.09 | -67.81 |
| C | BasketballDrill | 4.20 | -0.20 | -56.34 | -66.31 | -56.81 | -68.32 | -61.95 |
| BQMall | 2.21 | -0.13 | -66.71 | -66.51 | -66.06 | -67.92 | -66.80 |
| PartyScene | 1.45 | -0.11 | -59.45 | -59.31 | -59.52 | -59.79 | -59.52 |
| RaceHorses | 2.41 | -0.16 | -68.18 | -66.92 | -67.42 | -68.13 | -67.67 |
| D | BasketballPass | 2.43 | -0.14 | -50.72 | -51.82 | -66.57 | -68.06 | -59.30 |
| BQSquare | 1.61 | -0.14 | -60.05 | -59.73 | -59.57 | -60.27 | -59.90 |
| BlowingBubbles | 1.74 | -0.11 | -60.49 | -60.05 | -59.96 | -60.44 | -60.23 |
| RaceHorses | 1.96 | -0.13 | -65.51 | -64.33 | -64.21 | -63.75 | -64.45 |
| E | FourPeople | 3.14 | -0.18 | -64.33 | -72.74 | -72.54 | -73.49 | -70.77 |
| Johnny | 3.50 | -0.15 | -73.26 | -79.75 | -82.43 | -80.97 | -79.10 |
| KristenAndSara | 2.93 | -0.15 | -78.43 | -78.73 | -79.39 | -79.64 | -79.05 |
| Average | | 3.27 | -0.16 | -66.80 | -68.24 | -68.16 | -68.90 | -68.03 |

由表中数据可以看出，存在一个 BDBR、BDPSNR明显高于其余视频序列的数据，为了更直观地观察视频质量的变化，我们选取了表中BDPSNR损失最多的视频序列 Kimono对于该视频序列，我们将传统HM编码后的视频结果和本章提出的方法编码后的结果选取相同QP且同一帧进行对比，通过主观视觉图很难发现两者存在较为明显的区别，可以说明本章的方法虽然使BDPSNR有所升高，但是不影响整体的图像质量

### 消融实验

如上所述，本发明提出的多尺度CTU分区网络可以很好的学习到不同尺度下CU的特征信息，能够对CTU划分做出更加高效和准确的预测，从而降低HEVC的编码复杂度。 因此，为了能够准确评估本发明提出的多尺度CTU分区网络的有效性，我们改变三个网络的输入结构并命名为MSCNN-OneScale,它们都只考虑一级输入，对于预测CU是否需要划分的网络我们去掉了对尺度的输入，仅保留的尺度输入作为网络的学习对象，其余配置不变，同样，对于预测CU是否需要划分的网络我们去掉了对尺度的输入和尺度的输入, 保留的尺度输入作为网络的学习对象，其余配置不变。

表1 在100张CPIH测试集上CCS的有效性消融实验：CU、CU和CU的深度划分的平均准确率比较。

|  |  |  |
| --- | --- | --- |
| 方法 | MSCCN | MSCNN - OneScale |
| CU | 90.30% | 88.05% |
| CU | 87.55% | 86.51% |
| CU | 89.69% | 85.71% |

表2MSCNN的有效性消融实验

|  |  |  |
| --- | --- | --- |
| Model | 平均BD-BR | 平均 |
| MSCNN | 3.38% | 69.39% |
| MSCNN -OneScale | 11.5% | 63.3% |

为了验证我们提出的多尺度CTU分区网络能够学习到不同CU尺寸的图像特性，并对不同尺度的CU做出更加准确的划分预测，我们在CPIH数据集上进行CU、CU和CU的深度划分的平均准确率比较如表1所示。可以看出我们的多尺度CTU分区网络可以达到更准确的CU划分预测。

为了进一步验证我们的发明对HEVC复杂度降低上的效果，我们JCT-VC标准视频测试集上进行了测试如表2 所示，我们的发明在平均时间节省上比没有加入多尺度的方法高出了6.09%，而且BD-BR低了8.12，实验结果很好的体现的我们方法的优越性，在HEVC复杂度降低上，我们的发明具有明显的优势。

### 比较实验

表7示出了方法[1]、[2]、[3]、[4]、[5]和本发明提出的MCCN在JCT-VC所有18个标准视频测试集上的平均∆T比较，其中最好的性能已加粗标记

表7所示为方法[1]、[2]、[3]、[4]、[5]、和本发明提出的MCCN在JCT-VC所有的18个标准视频测试集上的平均BD-BR和平均比较。由表7可见，本发明提出的MSCCN具有最低的平均编码复杂度：仅以增加3.38%的平均BD-BR为代价，降低了69.39%的平均编码复杂度。特别的，首先本发明提出的MSCCN在平均方面均优于以上所有方法，与方法[1]、[2]、[3]、[4]和[5]相比，使用了MSCCN方法的编码复杂度实现了大幅的下降，平均分别提高了3.84%、8.73%、16.09%、7.54%和7.14%。因此，本发明提出的MSCCN明显优于现有的极具代表性的方法[1]、[2]、[3]、[4]和[5]。表7很好地证明了本发明提出的MSCCN的优越性。

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sequence | ref25 | | ref27 | | proposed | |  |  | Class | Sequence |
| BD-rate(%) | TS(%) | BD-rate(%) | TS(%) | BD-rate(%) | TS(%) |  |  |
| A | Traffic | 2.76 | -70.04 | 2.55 | -70.80 | 3.10 | -70.04 |  |  | A | Traffic |
| PeopleOnStreet | 2.16 | -62.98 | 2.37 | -61.00 | 2.40 | -69.73 |  |  | PeopleOnStreet |
| B | Kimono | 2.15 | -83.64 | 2.59 | -83.50 | 11.92 | -79.73 |  |  | B | Kimono |
| ParkScene | 2.13 | -70.79 | 1.96 | -67.50 | 2.38 | -70.73 |  |  | ParkScene |
| Cactus | 2.11 | -66.37 | 2.27 | -61.00 | 2.63 | -71.87 |  |  | Cactus |
| BasketballDrive | 4.60 | -75.78 | 4.27 | -76.30 | 6.99 | -65.87 |  |  | BasketballDrive |
| BQTerrace | 1.47 | -60.04 | 1.84 | -64.70 | 1.78 | -67.81 |  |  | BQTerrace |
| C | BasketballDrill | 2.97 | -58.13 | 2.86 | -53.00 | 4.20 | -61.95 |  |  | C | BasketballDrill |
| BQMall | 1.27 | -54.69 | 2.09 | -58.40 | 2.21 | -66.80 |  |  | BQMall |
| PartyScene | 0.50 | -40.17 | 0.66 | -44.50 | 1.45 | -59.52 |  |  | PartyScene |
| RaceHorses | 1.75 | -57.50 | 1.97 | -57.10 | 2.41 | -67.67 |  |  | RaceHorses |
| D | BasketballPass | 2.20 | -57.50 | 1.84 | -56.40 | 2.43 | -59.30 |  |  | D | BasketballPass |
| BQSquare | 0.19 | -46.99 | 0.91 | -45.80 | 1.61 | -59.90 |  |  | BQSquare |
| BlowingBubbles | 0.68 | -47.11 | 0.62 | -40.50 | 1.74 | -60.23 |  |  | BlowingBubbles |
| RaceHorses | 1.23 | -50.93 | 1.32 | -55.80 | 1.96 | -64.45 |  |  | RaceHorses |
| E | FourPeople | 3.12 | -67.49 | 3.11 | -71.30 | 3.14 | -70.77 |  |  | E | FourPeople |
| Johnny | 3.60 | -76.32 | 3.82 | -70.70 | 3.50 | -79.10 |  |  | Johnny |
| KristenAndSara | 3.21 | -74.01 | 3.46 | -74.80 | 2.93 | -79.05 |  |  | KristenAndSara |
| Average | | 2.12 | -62.25 | 2.25 | -61.84 | 3.27 | -68.03 |  |  | Average | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sequence | ref25 | | ref27 | | ref 33 | | ref 23 | | proposed | |
| BD-rate(%) | TS(%) | BD-rate(%) | TS(%) | BD-rate(%) | TS(%) | BD-rate(%) | TS(%) | BD-rate(%) | TS(%) |
| A | Traffic | 2.76 | -70.04 | 2.55 | -70.80 | 2.38 | -69.10 | 1.46 | -48.80 | 3.10 | -70.04 |
| PeopleOnStreet | 2.16 | -62.98 | 2.37 | -61.00 | 2.34 | -63.20 | 1.71 | -49.40 | 2.40 | -69.73 |
| B | Kimono | 2.15 | -83.64 | 2.59 | -83.50 | 1.73 | -78.40 | 1.46 | -47.70 | 11.92 | -79.73 |
| ParkScene | 2.13 | -70.79 | 1.96 | -67.50 | 1.70 | -64.80 | 1.54 | -49.50 | 2.38 | -70.73 |
| Cactus | 2.11 | -66.37 | 2.27 | -61.00 | 1.90 | -63.80 | 1.02 | -47.40 | 2.63 | -71.87 |
| BasketballDrive | 4.60 | -75.78 | 4.27 | -76.30 | 2.82 | -69.70 | 2.37 | -49.10 | 6.99 | -65.87 |
| BQTerrace | 1.47 | -60.04 | 1.84 | -64.70 | 1.35 | -57.20 | 0.82 | -46.70 | 1.78 | -67.81 |
| C | BasketballDrill | 2.97 | -58.13 | 2.86 | -53.00 | 2.41 | -53.30 | 1.48 | -47.00 | 4.20 | -61.95 |
| BQMall | 1.27 | -54.69 | 2.09 | -58.40 | 1.65 | -54.00 | 1.02 | -41.10 | 2.21 | -66.80 |
| PartyScene | 0.50 | -40.17 | 0.66 | -44.50 | 0.49 | -41.40 | 0.65 | -44.60 | 1.45 | -59.52 |
| RaceHorses | 1.75 | -57.50 | 1.97 | -57.10 | 1.52 | -56.20 | 0.85 | -48.70 | 2.41 | -67.67 |
| D | BasketballPass | 2.20 | -57.50 | 1.84 | -56.40 | 1.05 | -47.40 | 1.22 | -46.50 | 2.43 | -59.30 |
| BQSquare | 0.19 | -46.99 | 0.91 | -45.80 | 0.54 | -35.80 | 1.71 | -46.80 | 1.61 | -59.90 |
| BlowingBubbles | 0.68 | -47.11 | 0.62 | -40.50 | 0.32 | -31.80 | 1.03 | -44.20 | 1.74 | -60.23 |
| RaceHorses | 1.23 | -50.93 | 1.32 | -55.80 | 0.82 | -43.10 | 1.29 | -41.00 | 1.96 | -64.45 |
| E | FourPeople | 3.12 | -67.49 | 3.11 | -71.30 | 2.71 | -65.00 | 1.78 | -48.90 | 3.14 | -70.77 |
| Johnny | 3.60 | -76.32 | 3.82 | -70.70 | 3.16 | -75.20 | 2.22 | -49.90 | 3.50 | -79.10 |
| KristenAndSara | 3.21 | -74.01 | 3.46 | -74.80 | 2.68 | -71.40 | 2.21 | -49.50 | 2.93 | -79.05 |
| Average | | 2.12 | -62.25 | 2.25 | -61.84 | 1.75 | -57.82 | 1.44 | -47.04 | 3.27 | -68.03 |

## 本章小结

本章介绍了……

# 多尺度多输入互补分类网络



## 高效的CTU分区网络架构

## 研究展望

参考文献

1. Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on circuits and systems for video technology, 2012, 22(12): 1649-1668..
2. Wiegand T, Sullivan G J, Bjontegaard G, et al. Overview of the H. 264/AVC video coding standard[J]. IEEE Transactions on circuits and systems for video technology, 2003, 13(7): 560-576.
3. Lainema J, Bossen F, Han W J, et al. Intra coding of the HEVC standard[J]. IEEE transactions on circuits and systems for video technology, 2012, 22(12): 1792-1801.
4. Pourazad M T, Doutre C, Azimi M, et al. HEVC: The new gold standard for video compression: How does HEVC compare with H. 264/AVC?[J]. IEEE consumer electronics magazine, 2012, 1(3): 36-46.
5. Chen F, Jin D, Peng Z, et al. Fast intra coding algorithm for HEVC based on depth range prediction and mode reduction[J]. Multimedia Tools and Applications, 2018, 77(21): 28375-28394.
6. Cen Y F, Wang W L, Yao X W. A fast CU depth decision mechanism for HEVC[J]. Information Processing Letters, 2015, 115(9): 719-724
7. Xiong J, Li H, Wu Q, et al. A fast HEVC inter CU selection method based on pyramid motion divergence[J]. IEEE transactions on multimedia, 2013, 16(2): 559-564.
8. N. Kim, S. Jeon, H. Shim, B. Jeon, S.-C. Lim, H. Ko, Adaptive keypoint-based CU depth decision for HEVC intra coding,in: Proc. 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2016.
9. X.L. Shen, L. Yu, J. Chen, Fast coding unit size selection for HEVC based on Bayesian decision rule, in: Proc. 2012Picture Coding Symposium, 2012
10. X.J. Wang, Y.L. Xue, Fast HEVC inter prediction algorithm based on spatio-temporal block information, in: Proc. 2017IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2017.
11. X.G. Liu, Y.B. Liu, P.C. Wang, C.-F. Lai, H.-C. Chao, An Adaptive Mode Decision Algorithm Based on Video Texture Characteristics for HEVC Intra Prediction, IEEE Transactions on Circuits and Systems for Video Technology 27(8)(2017)
12. Shen L, Zhang Z, Liu Z. Effective CU size decision for HEVC intracoding[J]. IEEE Transactions on Image Processing, 2014, 23(10): 4232-4241.
13. Gweon R, Lee Y L. Early termination of CU encoding to reduce HEVC complexity[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2012, 95(7): 1215-1218.
14. Choi K, Park H M, Jang E S. JCTVC-F092 Coding tree pruning based CU early termination[J]. Joint Collaborative Team on Video Coding (JCTVC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Torino July, 2011: 1-11.
15. Yang J, Kim J, Won K, et al. Early SKIP detection for HEVC. JCT-VC of ITU-T SG16 WP3 and ISO[R]. IEC JTC1/SC29/WG11, Document JCTVC-G543, 2011.
16. Cho S, Kim M. Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(9): 1555-1564.
17. Zhang Y, Wang H, Li Z. Fast coding unit depth decision algorithm for interframe coding in HEVC[C]//2013 Data Compression Conference. IEEE, 2013: 53-62.
18. Zhao L, Fan X, Ma S, et al. Fast intra-encoding algorithm for high efficiency video coding[J]. Signal Processing: Image Communication, 2014, 29(9): 935-944
19. Khan M U K, Shafique M, Henkel J. An adaptive complexity reduction scheme with fast prediction unit decision for HEVC intra encoding[C]//2013 IEEE International Conference on Image Processing. IEEE, 2013: 1578-1582.
20. Zhang T, Sun M T, Zhao D, et al. Fast intra-mode and CU size decision for HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(8): 1714-1726.
21. Hu N, Yang E H. Fast mode selection for HEVC intra-frame coding with entropy coding refinement based on a transparent composite model[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(9): 1521-1532.
22. Liu D, Liu X, Li Y. Fast CU size decisions for HEVC intra frame coding based on support vector machines[C]//2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2016: 594-597.
23. Jamali M, Coulombe S. Fast HEVC intra mode decision based on RDO cost prediction[J]. IEEE transactions on broadcasting, 2018, 65(1): 109-122.
24. Yu X, Liu Z, Liu J, et al. VLSI friendly fast CU/PU mode decision for HEVC intra encoding: Leveraging convolution neural network[C]//2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015: 1285-1289.
25. Li T, Xu M, Deng X. A deep convolutional neural network approach for complexity reduction on intra-mode HEVC[C]//2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2017: 1255-1260.
26. Kim K, Ro W W. Fast CU depth decision for HEVC using neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(5): 1462-1473.
27. Xu M, Li T, Wang Z, et al. Reducing complexity of HEVC: A deep learning approach[J]. IEEE Transactions on Image Processing, 2018, 27(10): 5044-5059.
28. Kuanar S, Rao K R, Bilas M, et al. Adaptive CU mode selection in HEVC intra prediction: A deep learning approach[J]. Circuits, systems, and signal processing, 2019, 38: 5081-5102.
29. Li H, Wei G, Wang T, et al. Reducing Video Coding Complexity Based on CNN-CBAM in HEVC[J]. Applied Sciences, 2023, 13(18): 10135.
30. Qin L M, Zhu Z J, Bai Y Q, et al. A Complexity-Reducing HEVC Intra-Mode Method Based on VGGNet[J]. Journal of Computers, 2022, 33(4): 57-67
31. Fan J, Song L. Fast Intra-frame Prediction Algorithm for HEVC Based on Neural Networks and Adaptive Threshold[C]//Proceedings of the 2022 6th International Conference on Video and Image Processing. 2022: 127-134.
32. Wang T, Wei G, Li H, et al. A Method to Reduce the Intra-Frame Prediction Complexity of HEVC Based on D-CNN[J]. Electronics, 2023, 12(9): 2091.
33. Feng A, Gao C, Li L, et al. Cnn-based depth map prediction for fast block partitioning in hevc intra coding[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021: 1-6.
34. Hari P, Jadhav V, Rao B K N S. CTU Partition for Intra-Mode HEVC using Convolutional Neural Network[C]//2022 IEEE International Symposium on Smart Electronic Systems (iSES). IEEE, 2022: 548-551.
35. Lorkiewicz M, Stankiewicz O, Domanski M, et al. Fast Selection of INTRA CTU Partitioning in HEVC Encoders using Artificial Neural Networks[C]//2021 Signal Processing Symposium (SPSympo). IEEE, 2021: 177-182.
36. Ren W, Su J, Sun C, et al. An IBP-CNN based fast block partition for intra prediction[C]//2019 Picture Coding Symposium (PCS). IEEE, 2019: 1-5.
37. Feng Z, Liu P, Jia K, et al. HEVC fast intra coding based CTU depth range prediction[C]//2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2018: 551-555.
38. Li Y, Li L, Fang Y, et al. Bagged tree and ResNet-based joint end-to-end fast CTU partition decision algorithm for video intra coding[J]. Electronics, 2022, 11(8): 1264.
39. Imen W, Amna M, Fatma B, et al. Fast HEVC intra-CU decision partition algorithm with modified LeNet-5 and AlexNet[J]. Signal, Image and Video Processing, 2022, 16(7): 1811-1819.
40. Yao C, Xu C, Liu M. RDNet: Rate–Distortion-Based Coding Unit Partition Network for Intra-Prediction[J]. Electronics, 2022, 11(6): 916.
41. 王晓琰, 殷建芳, 王晓峰, 等. 关于连续出版会议论文著录格式的探讨[J]. 学报编辑论丛, 2019: 162-165.

**顺序编码制参考文献表的著录格式和示例**

**期刊论文示例:** [序号] 作者. 文题[J]. 期刊名, 年, 卷(期): 起-止页码.

1. WU D, YAN J, WANG H, et al. Social attribute aware incentive mechanism for device-to-device video distribution[J]. IEEE Transactions on Multimedia, 2017, 19(8): 1908-1920.
2. HEWITT J A. Technical services in 1983[J]. Library Resource Services, 1984, 28(3): 205-218.
3. 李炳穆. 理想的图书馆员和信息专家的素质与形象[J]. 图书情报工作, 2000(2): 5-8.

**会议论文示例:** [序号] 作者. 文题[C]. 会议名, 会议地, 会议年: 起-止页码.

1. BERGAMASCO F, ALBARELLI A, COSMO L, et al. Adopting an unconstrained ray model in light-field cameras for 3D shape reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3003-3012.
2. 中国力学学会. 第3届全国实验流体力学学术会议论文集[C]. 天津: 南开大学出版社, 1990: 10-12.

**普通图书示例:**  [序号] 作者. 书名[M]. 译者. 版本. 出版地: 出版者, 出版年: 起-止页码.

1. 竺可桢. 物理学[M]. 北京: 科学出版社, 1973: 56-60.
2. 汪继祥. 作者编辑手册[M]. 北京: 科学出版社, 2004.
3. 汪昂. (增补)本草备要[M]. 石印本. 上海: 同文书局, 1912: 31-32 .
4. ROOD H J. Logic and structured design for computer programmers[M]. 3rd ed. Watertown: Brooks/Cole Thomson Learning, 2001: 105-116.
5. CRAWFPRD W, GORMAN M. Future libraries: dreams, madness, & reality[M]. Chicago: American Library Association, 1995: 20-21.
6. 罗杰斯. 西方文明史：问题与源头[M]. 潘惠霞, 魏婧, 杨艳, 等译. 2版. 大连: 东北财经大学出版社, 2011: 15-16.

**学位论文示例:** [序号] 作者. 文题[D]. 授位单位所在地: 授位单位, 授位年: 起-止页码.

1. 李娜芬. 障碍环境中Swarm突现计算模型研究及行为控制[D]. 重庆: 重庆邮电大学, 2013: 11-47.

**报纸文章示例:** [序号] 作者. 文题[N]. 报纸名, 出版日期(版面数).

1. 顾春. 牢牢把握稳中求进的总基调[N]. 人民日报, 2012-03-31(3).
2. 丁文祥. 数字革命与竞争国际化[N]. 中国青年报, 2000-11-20(15).
3. 张田勤. 罪犯DNA库与生命伦理学计划[N]. 大众科技报, 2000-11-12(7).

**报告示例:** [序号] 作者. 文题[R]. 出版地: 出版者, 出版年.

1. 冯西桥. 核反应堆压力容器的LBB分析[R]. 北京: 清华大学核能技术设计研究院, 1997.
2. U. S. Department of Transportation Federal Highway Administration. Guidelines for handling excavated acid-producing material: PB 91-194001[R]. Springfield: U.S. Department of Commerce National Information Service, 1990.
3. World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.

**专利示例:** [序号] 发明人. 专利名: 专利号[P]. 授权日期.

1. 张凯军. 轨道火车及高速轨道火车紧急安全制动辅助装置: 201220158825. 2[P]. 2012-04-05.

**标准示例:** [序号] 发布单位. 标准名: 标准号[S]. 出版地: 出版者, 出版年: 起-止页码.

1. 国家技术监督局. 国际单位制及其应用: GB 3100-1993[S]. 北京: 中国标准出版社, 1994: 3-6.
2. 国家技术监督局. 有关量、单位和符号的一般规则: GB/T 3101-1993[S]. 北京: 中国标准出版社, 1994: 13-20.
3. 中国国家标准化管理委员会. 信息与文献 参考文献著录规则: GB/T 7714.1-2015[S]. 北京: 中国标准出版社, 2015: 1-18.

**电子文献示例：**[序号] 作者. 文题[文献类型标识/文献载体标识]. 出版地: 出版者, 出版年: 起-止页码 (更新或修改日期) [引用日期]. 获取或访问路径. 数字对象唯一标识符.

1. CLERC M. Discrete particle swarm optimization: a fuzzy combinatorial box[EB/OL]. 2010-07-16, http://clere.maurice.free.fr/pso/Fuzzy\_Discrere\_PSO/Fuzzy\_DPSO.htm.
2. Dublin core metadata elsment set:version 1.1[EB/OL].(2012-06-14)[2014-06-11].http://dublincore.org/docu-ments/dces/.
3. 河北绿洲生态环境科技有限公司. 一种荒漠化地区生态植被综合培育种植方法: 01129210.5[P/OL]. 2001-10-24[2002-05-28]. http://211.152.9.47/sipoasp/zlijs/hyjs-yx- new.asp? recid= 01129210.5&leixin.
4. 陈建军. 从数字地球到智慧地球[J/OL]. 国图资源导刊, 2010, 7(10): 93[2013-03-20].http://d.g.wanfangdata.com.cn/Periodical\_hunandz201010038.aspx.DOI:10.3969/j.issn.16725603.2010.10.038.
5. WALLS S C, BARICHIVICH W J, BROWN M E. Deluge and declines:the impact of precipitation extremes on amphibians in a changing climate[J/OL]. Biology, 2013, 2(1): 399-418[2013-11-04]. http://www.md-pi.com/2079-7737/2/1/399. DOI:10.3390/biology2010399.
6. 吴玉芳. 面向中文信息处理的现代汉语并列结构研究[D/OL]北京: 北京大学, 2003[2013-10-14].http://thesis.lib.pku.edu.cn/dlib/List.asp?lang= gb&type= Reader&DocGroupID= 4& DocID= 6328.

**著者-出版年制参考文献表的著录格式和示例**

顾春, 2012. 牢牢把握稳中求进的总基调[N]. 人民日报, 03-31 (3).

河北绿洲生态环境科技有限公司, 2001. 一种荒漠化地区生态植被综合培育种植方法: 01129210.5[P/OL]. http:// 211.152.9.47/sipoasp/zlijs/hyjs-yx-new.asp? recid= 01129210.5&leixin.

李炳穆, 2000. 理想的图书馆员和信息专家的素质与形象[J]. 图书情报工作, (2): 5-8.

尼葛洛庞帝, 1996.数字化生存[M]. 胡泳, 范海燕, 译. 海口: 海南出版社.

全国信息与文献标准化技术委员会, 2007. 学位论文编写规则: GB/T 7713.1-2006[S]. 北京: 中国标准出版社, 17-20.

全国信息与文献标准化技术委员会, 2015. 信息与文献 参考文献著录规则: GB/T 7714-2015[S]. 北京: 中国标准出版社, 1-18.

汪冰, 1997. 电子图书馆理论与实践研究[M]. 北京: 北京图书馆出版社: 16.

BAERGAMASCO F, ALBARELLI A, COSMO L, et al. 2015. Adopting an unconstrained ray model in light-field cameras for 3D shape reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 3003-3012.

DOWLER L, 1995. The research university’s dilemma:resource sharing and research in a transinstitutional environment[J]. Journal of Library Administration, 21(1/2): 5-26.

KENNEDY W J, GARRISON R E, 1975a. Morphology and genesis of nodular chalks and hardgrounds in the Upper Cretaceous of southern England[J]. Sedimentology, 22: 311-386.

KENNEDY W J, GARRISON R E, 1975b. Morphology and genesis of nodular phosphates in the cenomanian of South-east England[J]. Lathaia, 8: 339-360.

附录A 各学院中英文名称对照表

|  |  |  |
| --- | --- | --- |
| **序号** | **中文名称** | **英文名称** |
| 01 | 通信与信息工程学院 | School of Communications and Information Engineering |
| 02 | 计算机科学与技术学院/人工智能学院 | School of Computer Science and Technology / School of Artificial Intelligence |
| 03 | 自动化学院/工业物联网学院 | School of Automation / School of Industrial Internet of Things |
| 04 | 先进制造工程学院 | School of Advanced Manufacturing Engineering |
| 05 | 光电工程学院/重庆国际半导体学院 | School of Optoelectronic Engineering / Chongqing International Semiconductor College |
| … | …… | …… |

作者简介

## 基本情况

张某某，男，重庆人，1993年8月出生，重庆邮电大学XX学院XX专业2018级博士研究生。

## 教育和工作经历

2010.08～2014.06 重庆邮电大学光电工程学院，本科，专业：电子科学与技术

2014.08～2015.06 华为，技术研究工程师

2015.08～2018.06 重庆邮电大学光电工程学院，硕士研究生，专业：电子科学与技术

2018.08～2022.06 重庆邮电大学通信与信息工程学院，博士研究生，专业：信息与通信工程

## 攻读学位期间的研究成果

### 发表的学术论文和著作

1. **ZHANG M** , XX, XX. XXXX[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(9): 2543–2557. (SCI期刊)
2. XX, **张某某**, XX. XXXX [J]. 计算机学报, 2022. (已录用)
3. XX, XX, XX,**张某某**等. XXXX[M]. 科学出版社, 2021. (专著)

### 申请（授权）专利

1. **张某某**, XXX, XXX等. 专利名称: 专利号[P]. 授权日期.

### 参与的科研项目及获奖

格式：XXX项目, 项目名称, 起止时间, 完成情况, 作者贡献.

1. 国家自然科学基金重点项目, XXXX (No.000000), 2017.01-2020.12, 参与.
2. 重庆邮电大学博士研究生人才培养项目, XXXX (No.000000), 主持.
3. XXX, **张某某**, XXX等. 科研项目名称. 重庆市科技进步三等奖, 获奖日期.

致 谢

感谢老师、同学们的关心、支持和帮助！