

专利申请著录项目表



投诉二维码

接案号：TF240110129

(与： 同日申报)

(与： 一案两报)

客户联系人	姓名：唐述 电话：18223376355 电子邮箱：tangshu@cqupt.edu.cn		
	姓名：周广义 电话：15086765601 电子邮箱：1349604075@qq.com		
专利代理人	姓名：王世磊 电话：023-86898822-813；17265430177 电子邮箱：thy18@thycq.com		
特别提醒： 尊敬的客户，上述代理人的联系电话和邮箱为我司指定的工作电话和工作邮箱，为了保证您的技术方案的保密性和后续工作衔接的顺畅，请您务必采用上述联系方式与我方工作人员进行联系。			
①专利名称	一种高效的多尺度 CTU 分区网络架构		
②专利类型	发明 <input checked="" type="checkbox"/> 实用新型 <input type="checkbox"/> 外观设计 <input type="checkbox"/> 一案两报 <input type="checkbox"/>		是否费减： 是 <input checked="" type="checkbox"/> 否 <input type="checkbox"/>
③所有发明人	唐述、周广义、赵瑜、杨书丽、蒋忱迅、罗钧元、任川渝、韦哲韬、董文琦、梁雅琪、姚智皓		
④第一发明人国籍	中国	身份证号	510212198109200851
⑤申请人	申请人(1)	姓名或名称	重庆邮电大学
		组织结构代码/身份证号	125000004504018996
		邮政编码 400065	详细地址 重庆市南岸区黄桷垭崇文路 2 号
	申请人(2)	姓名或名称	
		组织结构代码/身份证号	
		邮政编码	详细地址
	申请人(3)	姓名或名称	
		组织结构代码/身份证号	
		邮政编码	详细地址
⑥提前公布	<input checked="" type="checkbox"/> 请求提前公布该专利申请（只适用于发明专利申请）		
⑦实质审查	<input checked="" type="checkbox"/> 在提交专利申请的同时提交实质审查请求（只适用于发明专利申请）		

特殊专利申请信息，涉及该内容时填写

⑧分案申请	原申请号： 针对的分案申请号： 原申请日： 年 月 日
⑨生物材料样品	保藏单位： 地址：
	保藏日期： 年 月 日 保藏编号： 分类命名：
	<input type="checkbox"/> 在提交专利申请的同时提交生物材料样品保藏及存活证明
⑩序列表	<input type="checkbox"/> 本专利申请涉及核苷酸或氨基酸序列表
⑪要求优先权声明	原受理结构名称： 在先申请日： 年 月 日 在先申请号：
	原受理结构名称： 在先申请日： 年 月 日 在先申请号：
备注	指定说明书附图中的图 1 为摘要附图

北京同恒源知识产权代理有限公司

一、本表由代理人预先填写，请联系人仔细核对信息是否正确，若有错误或缺漏请修改或补充。

二、本表第③栏，发明人是指对发明创造的实质性特点作出创造性贡献的人。发明人应当是个人（自然人）。发明人有两个发明人可以请求国家知识产权局不公布其姓名，若请求不公布姓名，应当在此栏所填写的相应发明人后面注明“（不公布姓名）”。

三、本表第④栏，第一发明人是香港、澳门、台湾地区，其国籍应填写为“中国”。第一发明人为中国内地居民的，还应当同时填写居民身份证号码（或军官证号码）。

四、本表第⑥栏，提前公布是指在发明专利申请初步审查合格后立即进入公布准备。如果不请求提前公布，则该发明专利申请将在自申请日起满十八个月时公布。**由于发明专利申请必须在公布以后才能进入实质审查程序，为了加快申请的审查进程，在申请人无特别要求的情况下，本公司默认勾选此栏。如申请人不要求提前公布，请去除勾选并及时通知代理人。**

五、本表第⑦栏，实质审查是指审查员对发明专利申请是否符合授权条件（包括新颖性、创造性、实用性、公开充分、单一性问题等）进行审查。申请人可以在自申请日（有优先权的，指优先权日）起三年内提出实质审查请求来启动实质审查程序。如果在提交专利申请的同时提交实质审查请求，则该发明专利申请在公布后立即进入实质审查阶段。**为了加快发明专利申请的审查进程，在申请人无特别要求的情况下，本公司默认勾选此栏。如申请人不想在提交专利申请的同时提交实质审查请求，请去除勾选并及时通知代理人。**

六、本表第⑨栏，发明专利申请涉及公众不能得到的生物材料的，应当填写此栏，并自申请日起四个月内提交由保藏单位出具的该生物材料样品的保藏及存活证明。

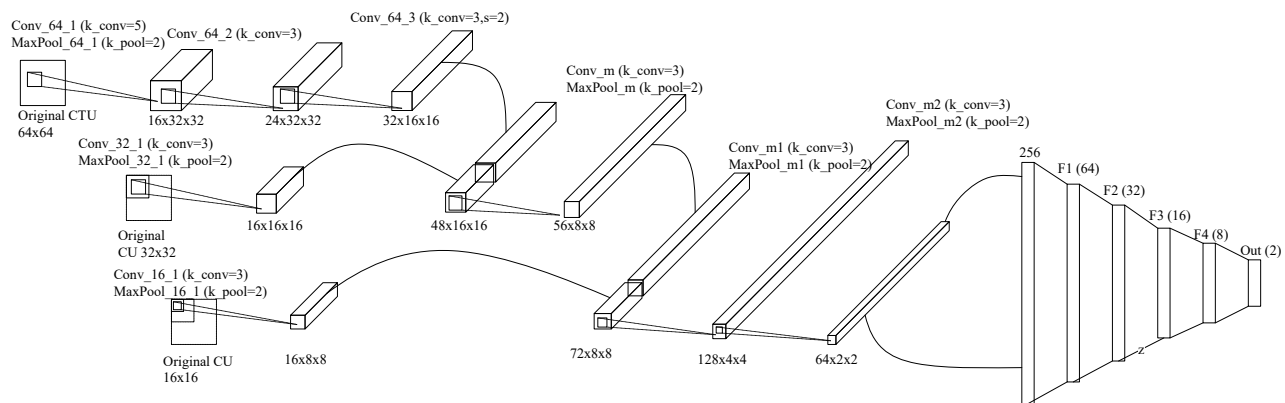
七、本表第⑩栏，发明专利申请涉及核苷酸或氨基酸序列表的，应当填写此栏，并在提交专利申请文件的同时提交核苷酸或氨基酸序列表的计算机可读形式副本。

八、本表第⑪栏，申请人要求外国或者本国优先权的，应当填写此栏。

说明书摘要

本发明涉及一种高效的多尺度 CTU 分区网络架构，属于视频编码领域。该网络架构包括一级预测结构、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 64×64 CU 是否需要划分，二级预测结构用于预测 32×32 CU 是否需要划分；三级预测结构用于预测 16×16 CU 是否需要划分；三级预测结构中包括多尺度特征提取层、深层特征提取层以及 CU 分区预测输出层，二级预测结构和一级预测结构中同样包括深层特征提取层以及 CU 分区预测输出层，而二级预测结构中是双尺度特征提取层，一级预测结构中是单尺度特征提取层。本发明对不同尺寸下 CU 的图像特征信息的提取，进而实现更加准确的 CTU 分区预测，从而降低 HEVC 的编码复杂度。

摘要附图



权利要求书

1、一种高效的多尺度 CTU 分区网络架构，其特征在于：其包括一级预测结构、二级预测结构以及三级预测结构，其中，所述一级预测结构用于预测 64×64 CU 是否需要划分，所述二级预测结构用于预测 32×32 CU 是否需要划分；所述三级预测结构用于预测 16×16 CU 是否需要划分；

所述一级预测结构中包括单尺度特征提取层、深层特征提取层以及 CU 分区预测输出层，其中，所述单尺度特征提取层对应输入 64×64 CU；

所述二级预测结构中包括双尺度特征提取层、深层特征提取层以及 CU 分区预测输出层，其中，所述双尺度特征包括第一层和第二层，第一层对应输入 64×64 CU，第二层对应输入 32×32 CU；

所述三级预测结构中包括多尺度特征提取层、深层特征提取层以及 CU 分区预测输出层，其中，所述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 64×64 CU，第二层对应输入 32×32 CU；第三层对应输入 16×16 CU。

2、根据权利要求 1 所述的一种高效的多尺度 CTU 分区网络架构，其特征在于：在所述一级预测结构中，执行如下操作：

1) 在一级预测结构的单尺度特征提取层中，采用 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息，并在该卷积层之后执行 1 次 2×2 的最大池化操作，得到一个 $16 \times 32 \times 32$ 的特征图；再采用 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，得到 $48 \times 16 \times 16$ 的特征图；再将 $48 \times 16 \times 16$ 特征图经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作，将特征图的分辨率降为 8×8 ，得到 $56 \times 8 \times 8$ 的特征图；

2) 进入一级预测结构的深层特征提取层，将 $56 \times 8 \times 8$ 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，分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图，其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图；

3) 在一级预测结构完成了基于卷积的特征提取和融合之后，采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果，5 个全连接层由 4 个隐藏层和 1 个输出层组成。

3、根据权利要求 1 所述的一种高效的多尺度 CTU 分区网络架构，其特征在于：在所述二级预测结构中，执行如下操作：

1) 在二级预测结构的多尺度特征提取层的第一层中, 针对 64×64 的 CU, 采用 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息, 并在该卷积层之后执行 1 次 2×2 的最大池化操作, 得到一个 $16 \times 32 \times 32$ 的特征图; 再采用 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, 得到 $32 \times 16 \times 16$ 的特征图;

2) 在二级预测结构的多尺度特征提取层的第二层中, 采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征, 在卷积层之后执行 1 次 2×2 的最大池化操作, 将第二级的特征图的分辨率降为 16×16 , 得到 $16 \times 16 \times 16$ 的特征图, 并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图; 然后, 将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作, 得到 $56 \times 8 \times 8$ 的特征图;

3) 然后进入二级预测结构的深层特征提取层, 将拼接后的 $56 \times 8 \times 8$ 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, 分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图, 其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图;

4) 在二级预测结构完成了基于卷积的二级特征提取和融合之后, 采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果, 5 个全连接层由 4 个隐藏层和 1 个输出层组成。

4、根据权利要求 1 所述的一种高效的多尺度 CTU 分区网络架构, 其特征在于: 在所述三级预测结构中, 执行如下操作:

1) 在三级预测结构的多尺度特征提取层的第一层中, 针对 64×64 的 CU, 采用 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息, 并在该卷积层之后执行 1 次 2×2 的最大池化操作, 得到一个 $16 \times 32 \times 32$ 的特征图; 再采用 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, 再次将特征图的分辨率降为 16×16 , 得到 $32 \times 16 \times 16$ 的特征图;

2) 在三级预测结构的多尺度特征提取层的第二层中, 采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征, 在卷积层之后执行 1 次 2×2 的最大池化操作, 将第二级的特征图的分辨率降为 16×16 , 得到 $16 \times 16 \times 16$ 的特征图, 并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图; 然后, 将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作, 得到 $56 \times 8 \times 8$ 的特征图;

3) 在三级预测结构的多尺度特征提取层的第三层中, 采用一个步长为 1 的 3×3 卷积层和一个 2×2 的最大池化层将输入的 16×16 CU 映射为 $16 \times 8 \times 8$ 的特征图, 并与第二级中的

56 × 8 × 8特征图再次进行通道维度上的拼接，得到72 × 8 × 8的特征图；

4) 然后进入三级预测结构的深层特征提取层，将再次拼接后的72 × 8 × 8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，分别得到维度为128 × 4 × 4和64 × 2 × 2的特征图，其中64 × 2 × 2特征图为最终得到的特征图；

5) 在三级预测结构完成了基于卷积的三级特征提取和融合之后，本发明采用 5 个连续的全连接层对64 × 2 × 2的特征图进行进一步的特征提取和输出最终的每个16 × 16CU 是否需要划分的二分类结果；5 个全连接层由 4 个隐藏层和 1 个输出层组成。

5、根据权利要求 1-4 任一所述的一种高效的多尺度 CTU 分区网络架构，其特征在于：在所述一级预测结构、所述二级预测结构以及所述三级预测结构的基于卷积的特征提取过程中，每个卷积层均采用 PReLU 作为非线性激活函数。

6、根据权利要求 1-4 任一所述的一种高效的多尺度 CTU 分区网络架构，其特征在于：在所述一级预测结构、所述二级预测结构以及所述三级预测结构中均采用交叉熵作为其损失函数，如公式(1)所示：

$$L = -\frac{1}{N} \sum_{n=1}^N [y_i^n \log(p_i^n)] \quad (1)$$

在公式(1)中，N 表示不同尺度下 CU 的数量，在一级预测结构中 N=1，在二级预测结构中 N=4，在三级预测结构中 N=16；

y_i^n 和 p_i^n 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， $i \in [0,1]$ 。

一种高效的多尺度 CTU 分区网络架构

技术领域

本发明属于视频编码技术领域，涉及一种高效的多尺度 CTU 分区网络架构。

背景技术

近年来，降低 HEVC 的编码复杂度的方法主要分为两大类：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）。

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。早期，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 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 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。Kim 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 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。在这些方法中都试图寻找到一种影响 CTU 划分的图像特征，以此作为 CTU 分区的依据。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。

近年来，深度学习神经网络，尤其是 CNN，因为其强大的特征提取能力和非线性表达能力，人们已将其用于 CTU 的深度划分中，出现了基于深度学习的方法。Yu 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，并且对所有深度的 CU 采用相同的网络架构。Li 等人针对不同深度的 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。Xu 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。Feng 等人提出了一种基于 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。Ren 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CTU块的深度划分

由以上分析可知, 虽然现有的基于深度学习的预测方法能够通过自主学习CTU的结构来做出对应的CTU深度划分决策, 但是, 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。由于缺乏对小尺度CU的特征提取和融合多尺度CU特性信息, 导致这些网络不能很好的进行更准确CTU划分, 因此现有的基于学习的方法仍有较大的提升空间。

在现有的基于学习的方法中, 大家普遍的使用以CNN为基础构建的神经网络, 并将整块CTU的亮度分量作为神经网络学习的数据进行训练, 在这些方法中大多只考虑改变神经网络神经深度或者卷积核的尺寸来提高对CTU的特征提取能力, 但却忽略了CTU中不同尺度下的CU之间的联系。

发明内容

有鉴于此, 本发明的目的在于提供一种高效的多尺度 CTU 分区网络架构。

为达到上述目的, 本发明提供如下技术方案:

一种高效的多尺度 CTU 分区网络架构, 其包括一级预测结构、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 64×64 CU 是否需要划分, 二级预测结构用于预测 32×32 CU 是否需要划分; 三级预测结构用于预测 16×16 CU 是否需要划分;

一级预测结构中包括单尺度特征提取层、深层特征提取层以及 CU 分区预测输出层, 其中, 单尺度特征提取层对应输入 64×64 CU;

二级预测结构中包括双尺度特征提取层、深层特征提取层以及 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 64×64 CU, 第二层对应输入 32×32 CU;

三级预测结构中包括多尺度特征提取层、深层特征提取层以及 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 64×64 CU, 第二层对应输入 32×32 CU; 第三层对应输入 16×16 CU。

进一步, 在一级预测结构中, 执行如下操作:

1) 在一级预测结构的单尺度特征提取层中, 采用 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息, 并在该卷积层之后执行 1 次 2×2 的最大池化操作, 得到一个 $16 \times 32 \times 32$ 的特征图; 再采用 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, 得到 $48 \times 16 \times 16$ 的特征图; 再将 $48 \times 16 \times 16$ 特征图经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作, 将特征图的分辨率降为 8×8 , 得到

$56 \times 8 \times 8$ 的特征图；

2) 进入一级预测结构的深层特征提取层，将 $56 \times 8 \times 8$ 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，分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图，其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图；

3) 在一级预测结构完成了基于卷积的特征提取和融合之后，采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果，5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步，在二级预测结构中，执行如下操作：

1) 在二级预测结构的多尺度特征提取层的第一层中，针对 64×64 的 CU，采用了 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息，并在该卷积层之后执行 1 次 2×2 的最大池化操作，得到一个 $16 \times 32 \times 32$ 的特征图；再采用 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，得到 $32 \times 16 \times 16$ 的特征图；

2) 在二级预测结构的多尺度特征提取层的第二层中，采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征，在卷积层之后执行 1 次 2×2 的最大池化操作，将第二级的特征图的分辨率降为 16×16 ，得到 $16 \times 16 \times 16$ 的特征图，并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图；然后，将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作，得到 $56 \times 8 \times 8$ 的特征图；

3) 然后进入二级预测结构的深层特征提取层，将拼接后的 $56 \times 8 \times 8$ 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，分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图，其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图；

4) 在二级预测结构完成了基于卷积的二级特征提取和融合之后，采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果，5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步，在三级预测结构中，执行如下操作：

1) 在三级预测结构的多尺度特征提取层的第一层中，针对 64×64 的 CU，采用了 1 个步长为 1 的 5×5 卷积层来提取其中的特征信息，并在该卷积层之后执行 1 次 2×2 的最大池化操作，得到一个 $16 \times 32 \times 32$ 的特征图；再采用 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，再次将特征图的分辨率降为 16×16 ，得到 $32 \times 16 \times 16$ 的特征图；

2) 在三级预测结构的多尺度特征提取层的第二层中, 采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征, 在卷积层之后执行 1 次 2×2 的最大池化操作, 将第二级的特征图的分辨率降为 16×16 , 得到 $16 \times 16 \times 16$ 的特征图, 并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图; 然后, 将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作, 得到 $56 \times 8 \times 8$ 的特征图;

3) 在三级预测结构的多尺度特征提取层的第三层中, 采用一个步长为 1 的 3×3 卷积层和一个 2×2 的最大池化层将输入的 16×16 CU 映射为 $16 \times 8 \times 8$ 的特征图, 并与第二级中的 $56 \times 8 \times 8$ 特征图再次进行通道维度上的拼接, 得到 $72 \times 8 \times 8$ 的特征图;

4) 然后进入三级预测结构的深层特征提取层, 将再次拼接后的 $72 \times 8 \times 8$ 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, 分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图, 其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图;

5) 在三级预测结构完成了基于卷积的三级特征提取和融合之后, 本发明采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果; 5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步, 在一级预测结构、二级预测结构以及三级预测结构的基于卷积的特征提取过程中, 每个卷积层均采用 PReLU 来作为非线性激活函数。

进一步, 在一级预测结构、二级预测结构以及三级预测结构中均采用交叉熵作为其损失函数, 如公式(1)所示:

$$L = -\frac{1}{N} \sum_{n=1}^N [y_i^n \log(p_i^n)] \quad (1)$$

在公式(1)中, N 表示不同尺度下 CU 的数量, 在一级预测结构中 N=1; 在二级预测结构中 N=4; 在三级预测结构中 N=16;

y_i^n 和 p_i^n 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, $i \in [0,1]$ 。

本发明的有益效果在于: 提出了一种新的多尺度的卷积神经网络来实现对不同尺寸下 CU 的图像特征信息的提取, 进而实现更加准确的 CTU 分区预测, 从而降低 HEVC 的编码复杂度。实验结果表明, 与近几年极具代表神经网络方法相比, 本发明提出的多尺度的卷积神经网络, 可以更加高效的预测 CTU 分区信息有效的降低了 HEVC 的编码复杂度。

本发明的其他优点、目标和特征在某种程度上将在随后的说明书中进行阐述, 并且在某种程度上, 基于对下文的考察研究对本领域技术人员而言将是显而易见的, 或者可以从本发明的实践中得到教导。本发明的目标和其他优点可以通过下面的说明书来实现和获得。

附图说明

为了使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明作优选的详细描述，其中：

图 1 为本发明的三级预测结构的网络结构图；

图 2 为本发明的二级预测结构的网络结构图；

图 3 为本发明的一级预测结构的网络结构图。

具体实施方式

以下通过特定的具体实例说明本发明的实施方式，本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用，本说明书中的各项细节也可以基于不同观点与应用，在没有背离本发明的精神下进行各种修饰或改变。需要说明的是，以下实施例中所提供的图示仅以示意方式说明本发明的基本构想，在不冲突的情况下，以下实施例及实施例中的特征可以相互组合。

其中，附图仅用于示例性说明，表示的仅是示意图，而非实物图，不能理解为对本发明的限制；为了更好地说明本发明的实施例，附图某些部件会有省略、放大或缩小，并不代表实际产品的尺寸；对本领域技术人员来说，附图中某些公知结构及其说明可能省略是可以理解的。

本发明实施例的附图中相同或相似的标号对应相同或相似的部件；在本发明的描述中，需要理解的是，若有术语“上”、“下”、“左”、“右”、“前”、“后”等指示的方位或位置关系为基于附图所示的方位或位置关系，仅是为了便于描述本发明和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作，因此附图中描述位置关系的用语仅用于示例性说明，不能理解为对本发明的限制，对于本领域的普通技术人员而言，可以根据具体情况理解上述术语的具体含义。

请参阅图 1～图 3，为一种高效的多尺度 CTU 分区网络架构。

目前基于深度学习的方法在 HEVC 帧内编码复杂度降低上仍然存在一些不足：现有的一些基于深度学习的方法几乎都将整块 CTU 的图像数据作为神经网络学习的对象，没有考虑不同尺度 CU 之间的特征关联导致网络很难对小尺寸的 CU 做出更加准确的划分，因此现有的基于深度学习的方法仍有较大的提升空间。

针对现有方法存在的缺陷，本发明提出了一种新的多尺度的卷积神经网络(Multi-scale Convolutional Neural Networks)来实现对不同尺寸下 CU 的图像特征信息的提取，进而实现更

加准确的 CTU 分区预测。从而降低 HEVC 的编码复杂度。最后，本发明做了大量的实验，实验结果表明，与近几年极具代表神经网络方法相比，本发明提出的多尺度的卷积神经网络，可以更加高效的预测 CTU 分区信息有效的降低了 HEVC 的编码复杂度。

实施例 1

本实施例对本发明的多尺度 CTU 分区网络架构进行详细介绍。

在 HEVC 的帧内编码中规定：编码树单元（CTU）的大小为 64×64 ，而 CTU 中编码单元（CU）的大小则根据其深度的不同分别为 64×64 、 32×32 、 16×16 和 8×8 ，HEVC 会逐层判读当前 CU 是否需要划分，如果 64×64 的 CU 需要划分则划分为 4 个 32×32 的 CU，如果其中 32×32 的 CU 也需要划分则将该 32×32 的 CU 划分为 4 个 16×16 的 CU，最后如果 16×16 的 CU 也需要划分则继续划分为 8×8 的 CU，本发明根据 HEVC 的 CU 划分过程设计了一种用于判断不同 CU 是否需要划分的多尺度卷积神经网络。本发明提出的多尺度卷积神经网络至少包括一级预测结构、二级预测结构以及三级预测结构，分别用于预测 64×64 、 32×32 、 16×16 的 CU 是否需要划分，由于 HEVC 中 8×8 的 CU 是最小的分区单元，故不需要为其创建网络。

本发明提出的用于预测 16×16 CU 是否需要划分的三级预测结构如图 1 所示，三级预测结构主要包括三个部分：（1）多尺度特征提取层；（2）深层特征提取层；（3）CU 分区预测输出层。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 64×64 的 CU，第二层对应 32×32 的 CU，第三层对应 16×16 的 CU；将 64×64 的 CU 输入到第一层，同时将该 64×64 CU 中所有的 32×32 CU 和所有的 16×16 CU 分别对应到第二层和第三层的输入。

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 16×16 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：

1) 在三级预测结构的多尺度特征提取层的第一层中，针对 64×64 的 CU，本发明首先采用了 1 个步长为 1 的 5×5 卷积层 ($k_{\text{conv}}=5$) 来提取其中的特征信息，并在该卷积层之后执行 1 次 2×2 的最大池化操作 ($k_{\text{pool}}=2$)，在保留显著特征的同时实现降维，将特征图的分辨率降为 32×32 ，得到一个 $16 \times 32 \times 32$ 的特征图；再采用 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，再次将特征图的分辨率降为 16×16 ，得到 $32 \times 16 \times 16$ 的特征图；

2) 在三级预测结构的多尺度特征提取层的第二层中，本发明首先采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征，并且同样在卷积层之后执行 1 次 2×2 的最大池化操作，将第二级的特征图的分辨率降为 16×16 ，得到 $16 \times 16 \times 16$ 的特征图，并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图；然后，将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作，将拼接特征图的分辨率降为 8×8 ，得到 $56 \times 8 \times 8$ 的特征图；

3) 在三级预测结构的多尺度特征提取层的第三层中，本发明首先采用一个步长为 1 的 3×3 卷积层和一个 2×2 的最大池化层将输入的 16×16 CU 映射为 $16 \times 8 \times 8$ 的特征图，并与第二级中的 $56 \times 8 \times 8$ 特征图再次进行通道维度上的拼接，得到 $72 \times 8 \times 8$ 的特征图；

4) 然后进入三级预测结构的深层特征提取层，将再次拼接后的 $72 \times 8 \times 8$ 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，分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图，其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图；

5) 在三级预测结构完成了基于卷积的三级特征提取和融合之后，本发明采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果。5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步地，对于预测 32×32 的 CU 是否需要划分的二级预测结构如图 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；与三级预测结构不同的是在多尺度上面，二级网络结构取消了 16×16 尺度的输入，保留了 64×64 、 32×32 的两个尺度，即二级预测结构的多尺度特征层包括两层，其余，第一层对应 64×64 的 CU，第二层对应 32×32 的 CU。因为二级预测结构是用于预测 32×32 尺寸的 CU 是否需要划分，所以二级预测结构只需要学习 32×32 尺度的 CU 特性就可以了。表 2 示出了用于预测 32×32 CU 是否需要划分的二级预测结构的网络配置参数。

表 2

Layer	Filter Size	Number of Filters
-------	-------------	-------------------

说明书

Conv-64-1	5x5	16
Conv-64-2	3x3	24
Conv-64-3	3x3	48
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

参照表 2，对于二级预测结构而言，包括如下操作：

1) 在二级预测结构的双尺度特征提取层的第一层中，针对 64×64 的 CU，本发明首先采用了 1 个步长为 1 的 5×5 卷积层 ($k_{\text{conv}}=5$) 来提取其中的特征信息，并在该卷积层之后执行 1 次 2×2 的最大池化操作 ($k_{\text{pool}}=2$)，在保留显著特征的同时实现降维，将特征图的分辨率降为 32×32 ，得到一个 $16 \times 32 \times 32$ 的特征图；再采用 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，再次将特征图的分辨率降为 16×16 ，得到 $32 \times 16 \times 16$ 的特征图；

2) 在二级预测结构的双尺度特征提取层的第二层中，本发明首先采用 1 个步长为 1 的 3×3 的卷积层来提取 32×32 CU 的特征，并且同样在卷积层之后执行 1 次 2×2 的最大池化操作，将第二级的特征图的分辨率降为 16×16 ，得到 $16 \times 16 \times 16$ 的特征图，并将其与第一级的 $32 \times 16 \times 16$ 特征图在通道维度上进行拼接得到 $48 \times 16 \times 16$ 特征图；然后，将拼接后的 $48 \times 16 \times 16$ 特征图再经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作，将拼接特征图的分辨率降为 8×8 ，得到 $56 \times 8 \times 8$ 的特征图；

3) 然后进入二级预测结构的深层特征提取层，将拼接后的 $56 \times 8 \times 8$ 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，分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图，其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图；

4) 在二级预测结构完成了基于卷积的二级特征提取和融合之后，本发明采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果。5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步地，本发明的预测 64×64 的 CU 是否需要划分的一级预测结构如图 3 所示，与前两个三级预测结构和二级预测结构不同的是，一级预测结构仅输入了 64×64 的尺度提供给网络进行学习，即一级预测结构包括单尺度特征提取层、深层特征提取层以及 CU 分区预测输出，其中，单尺度特征提取层作为一级预测结构的输入，其对应仅输入 64×64 的尺度的 CU。

表 3

Layer	Filter Size	Number of Filters
Conv-64-1	5x5	16
Conv-64-2	3x3	24
Conv-64-3	3x3	32
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 3 示出的用于预测 64×64 是否划分的一级预测结构的网络配置参数,对于一级预测结构,包括如下操作:

1) 在一级预测结构的单尺度特征提取层中,针对 64×64 的 CU,本发明首先采用了 1 个步长为 1 的 5×5 卷积层($k_{\text{conv}}=5$)来提取其中的特征信息,并在该卷积层之后执行 1 次 2×2 的最大池化操作 ($k_{\text{pool}}=2$),在保留显著特征的同时实现降维,将特征图的分辨率降为 32×32 ,得到一个 $16 \times 32 \times 32$ 的特征图;再采用 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,再次将特征图的分辨率降为 16×16 ,得到 $48 \times 16 \times 16$ 的特征图;再将 $48 \times 16 \times 16$ 特征图经过 1 个 3×3 的卷积层和 1 个 2×2 的最大池化操作,将特征图的分辨率降为 8×8 ,得到 $56 \times 8 \times 8$ 的特征图;

2) 然后进入一级预测结构的深层特征提取层,将 $56 \times 8 \times 8$ 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,分别得到维度为 $128 \times 4 \times 4$ 和 $64 \times 2 \times 2$ 的特征图,其中 $64 \times 2 \times 2$ 特征图为最终得到的特征图;

3) 在一级预测结构完成了基于卷积的特征提取和融合之后,本发明采用 5 个连续的全连接层对 $64 \times 2 \times 2$ 的特征图进行进一步的特征提取和输出最终的每个 16×16 CU 是否需要划分的二分类结果。5 个全连接层由 4 个隐藏层和 1 个输出层组成。

进一步,在以上基于卷积的特征提取过程中,每个卷积层都采用 PReLU 来作为非线性激活函数。

进一步,在这三个网络中本发明都采用交叉熵作为其损失函数,如公式(1)所示:

$$L = -\frac{1}{N} \sum_{n=1}^N [y_i^n \log(p_i^n)] \quad (1)$$

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 64×64 的 CU 是否需要划分,所以此时 $N=1$, 64×64 的 CU 又包含 4 个 32×32 的 CU,所以在图 2

的网络中 $N=4$ ，每个 32×32 的 CU 又包含 4 个 16×16 的 CU 所以在图 1 中 $N=16$ 。 y_i^n 和 p_i^n 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，由于本发明是需要预测每个 CU 是否需要划分，是一个二分类问题，所以 $i \in [0,1]$ 。

实施例 2

本实施例把在 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，本实施例采用 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 (Bjontegaard 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 ΔT ，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：

$$\Delta T = \frac{T_{HM} - T_{test}}{T_{HM}} \quad (2)$$

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)，动量设置为 0.9，初始化学率为 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。

本实施针对本发明的方法进行了消融实验。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，本实施例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 32×32 CU 是否需要划分的网络本实施例去掉了 64×64 尺度的输入，仅保留 32×32 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 16×16 CU 是否需要划分的网络本实施例去掉了 64×64 尺度的输入和 32×32 尺度的输入，保留

说明书

16 × 16 的尺度输入作为网络的学习对象，其余配置不变。在 100 张 CPIH 测试集上 CCS 的有效性消融实验，表 4 示出了 64 × 64CU、32 × 32CU 和 16 × 16CU 的深度划分的平均准确率比较结果；表 5 示出了 MSCNN 的有效性消融实验结果。

表 4

方法	MSCCN	MSCNN-OneScale
64 × 64CU	90.30%	88.05%
32 × 32CU	87.55%	86.51%
16 × 16CU	89.69%	85.71%

表 5

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

为了验证本发明提出的多尺度 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，本实施例在 CPIH 数据集上进行 64 × 64CU、32 × 32CU 和 16 × 16CU 的深度划分的平均准确率比较如表 4 所示。可以看出本发明的多尺度 CTU 分区网络可以达到更准确的 CU 划分预测。

为了进一步验证本发明对 HEVC 复杂度降低上的效果，本实施例 JCT-VC 标准视频测试集上进行了测试如表 5 所示，本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本发明的方法的优越性，在 HEVC 复杂度降低上，本发明具有明显的优势。

为了证明本发明提出的方法的优越性，本实施例将和几个极具代表性的方法进行了比较，比较方法如表 6 所示：

表 6

文献标号	对比文献
[1]	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.
[2]	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.
[3]	Jamali M, Coulombe S. Fast HEVC intra mode decision based on RDO cost prediction[J]. IEEE transactions on broadcasting, 2018, 65(1): 109-122
[4]	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.
[5]	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.

说明书

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 ΔT 比较，其中最好的性能已加粗标记。

表 7						
方法	Proposed	[1]	[2]	[3]	[4]	[5]
ΔT	69.39%	65.55%	60.63%	47.30%	61.85%	62.25%

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 ΔT 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 ΔT 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 ΔT 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。

最后说明的是，以上实施例仅用以说明本发明的技术方案而非限制，尽管参照较佳实施例对本发明进行了详细说明，本领域的普通技术人员应当理解，可以对本发明的技术方案进行修改或者等同替换，而不脱离本技术方案的宗旨和范围，其均应涵盖在本发明的权利要求范围当中。

说明书附图

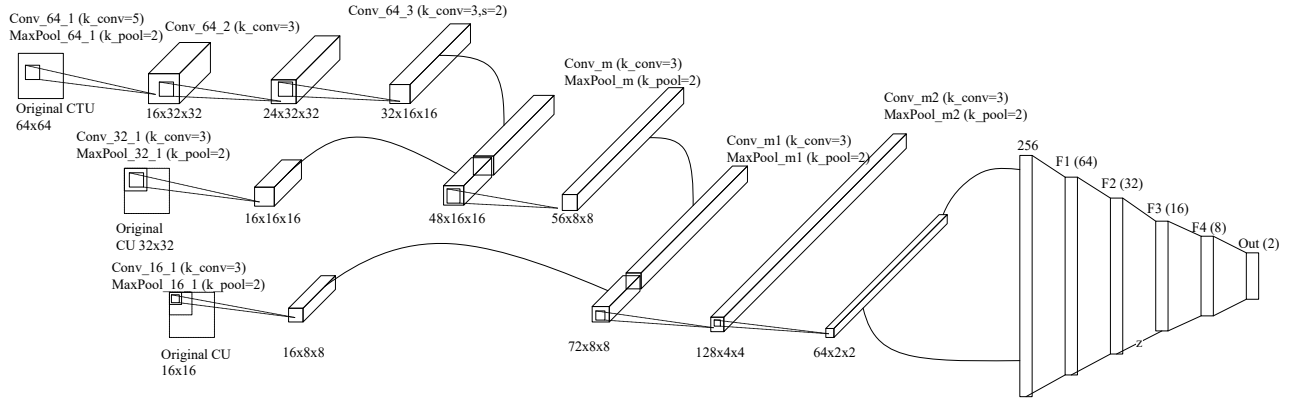


图 1

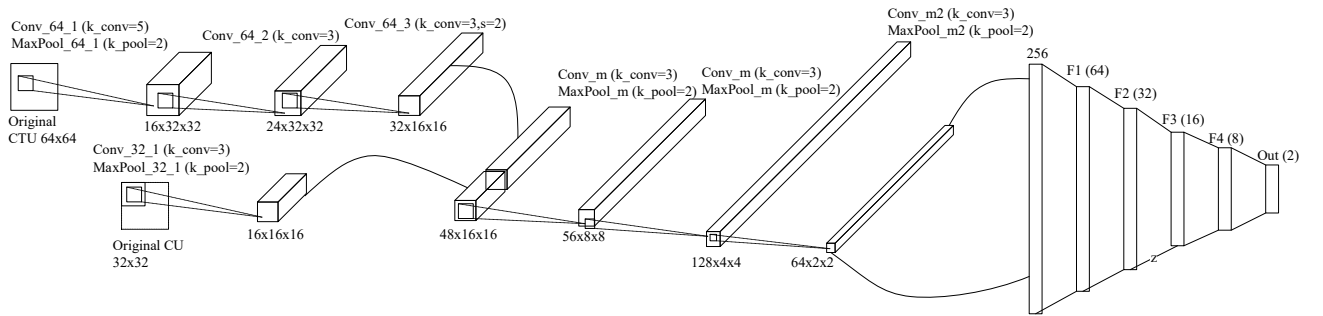


图 2

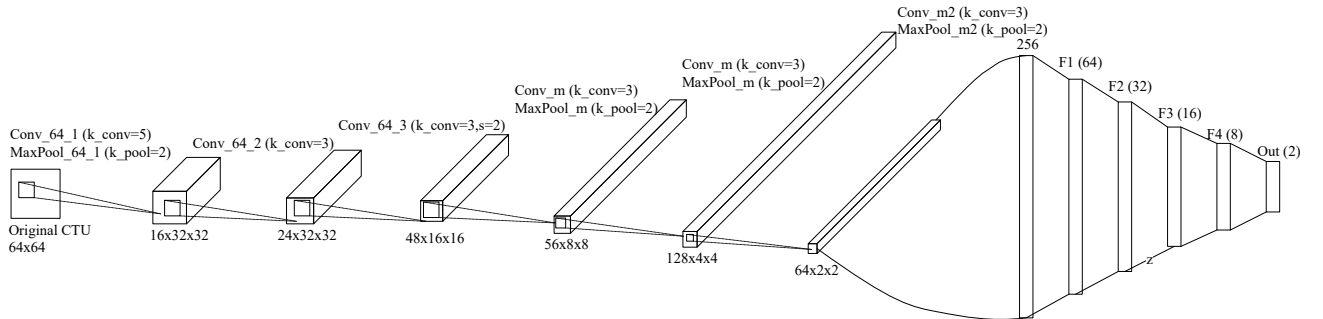


图 3