Zig-zag Scanning 锯齿形扫描

(a) Quantized coefficients

Linearized vector

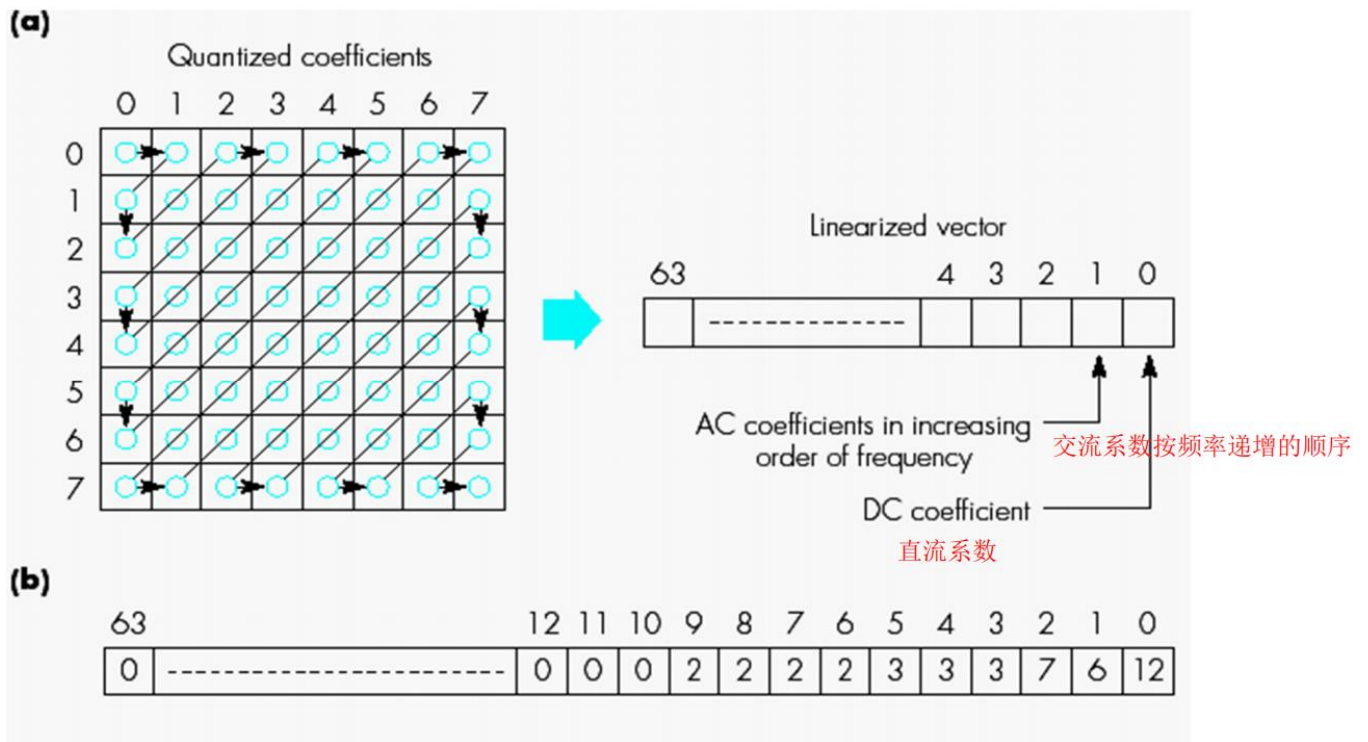AC coefficients in increasing order of frequency 交流系数按频率递增的顺序
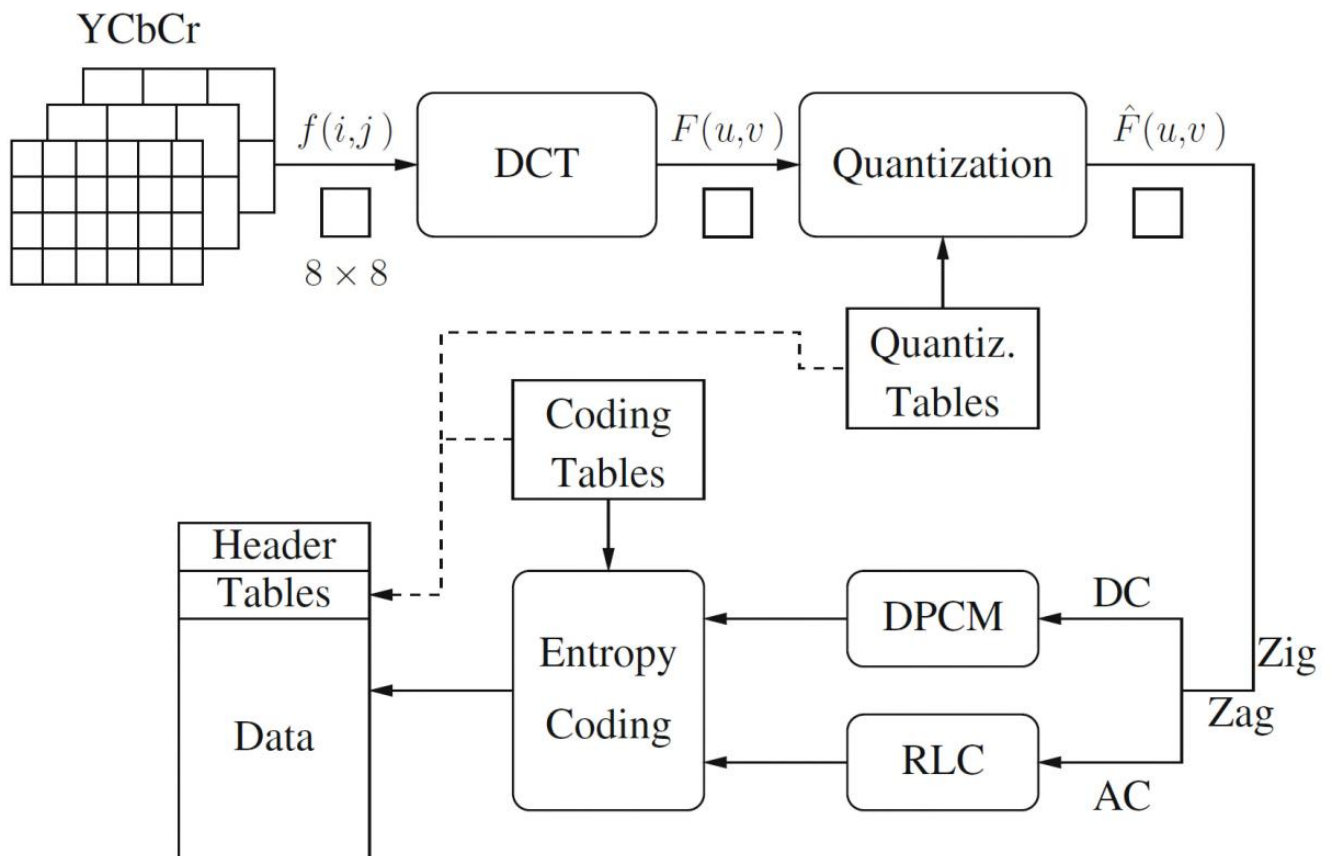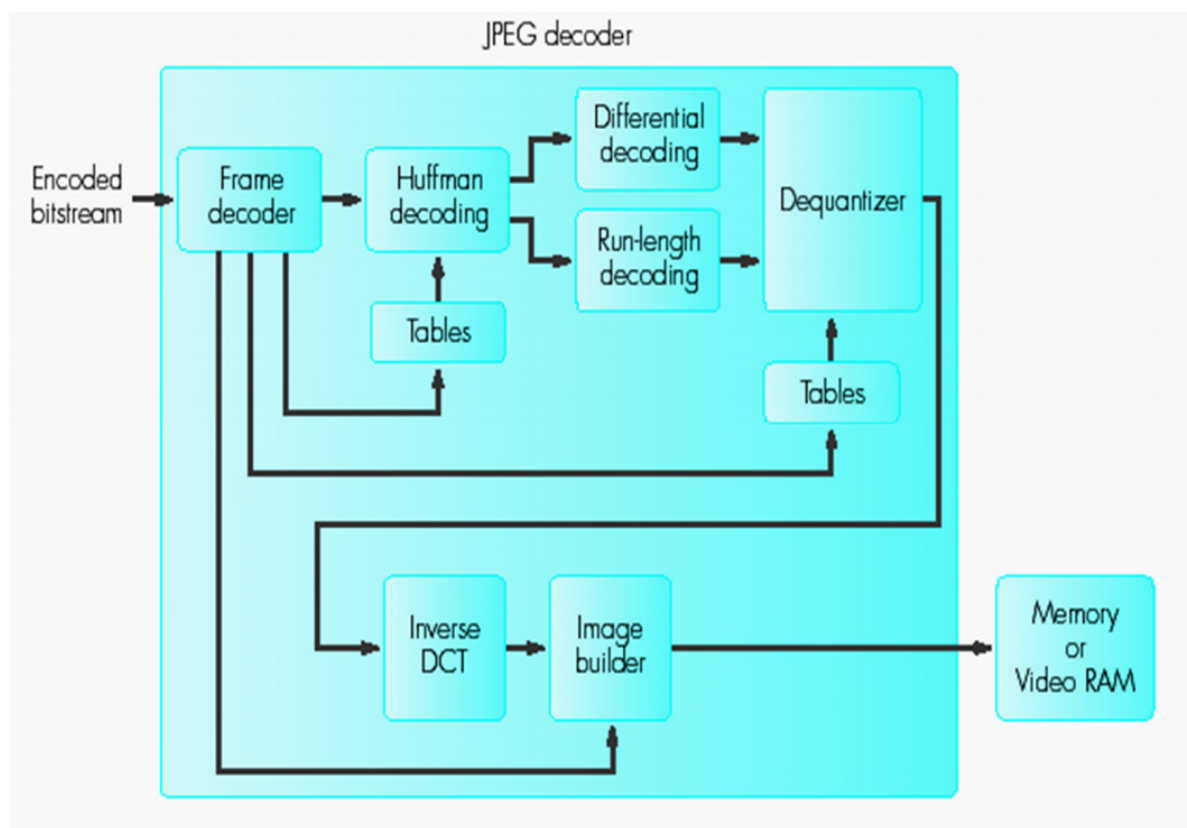
DC coefficient 直流系数

(b)

JPEG Encoder

Fig. 9.1: Block diagram for JPEG encoder.

Source: Ze-Nian Li, Mark S. Drew, Jiangchuan Liu, Fundamental of Multimedia, Springer 2021

JPEG Decoder

JPEG decoder

Differential decoding

Encoded bitstream

Frame decoder

Huffman decoding

Run-length decoding

Dequantizer

Tables

Tables

Inverse DCT

Image builder

Memory or Video RAM

Hierarchical Search

分层搜索；阶层式搜寻

- A three-level hierarchical search, the original image is at Level 0.

- Images at Levels 1 and 2 are obtained by down-sampling from the previous levels by a factor of 2, and the initial search is conducted at Level 2.
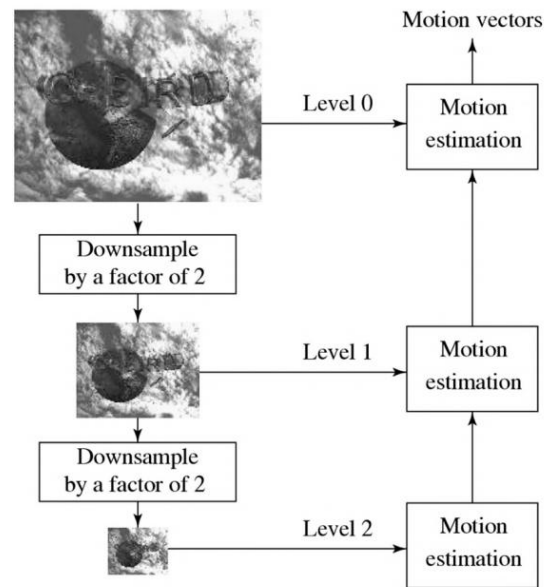
三级分层搜索，原始图像处于o级。

1级和2级的图像是通过对前两级进行2倍的降采样获得的，初始搜索在2级进行。

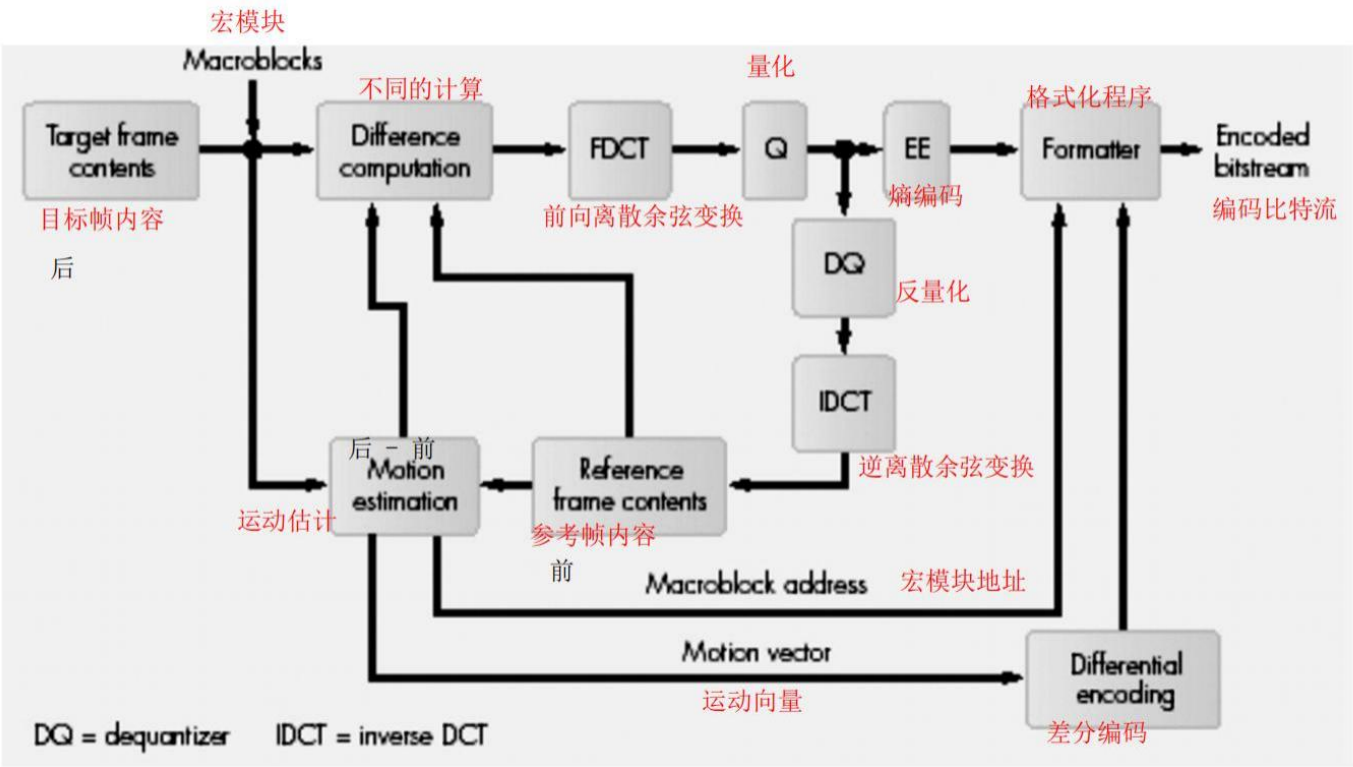A Three-level Hierarchical Search
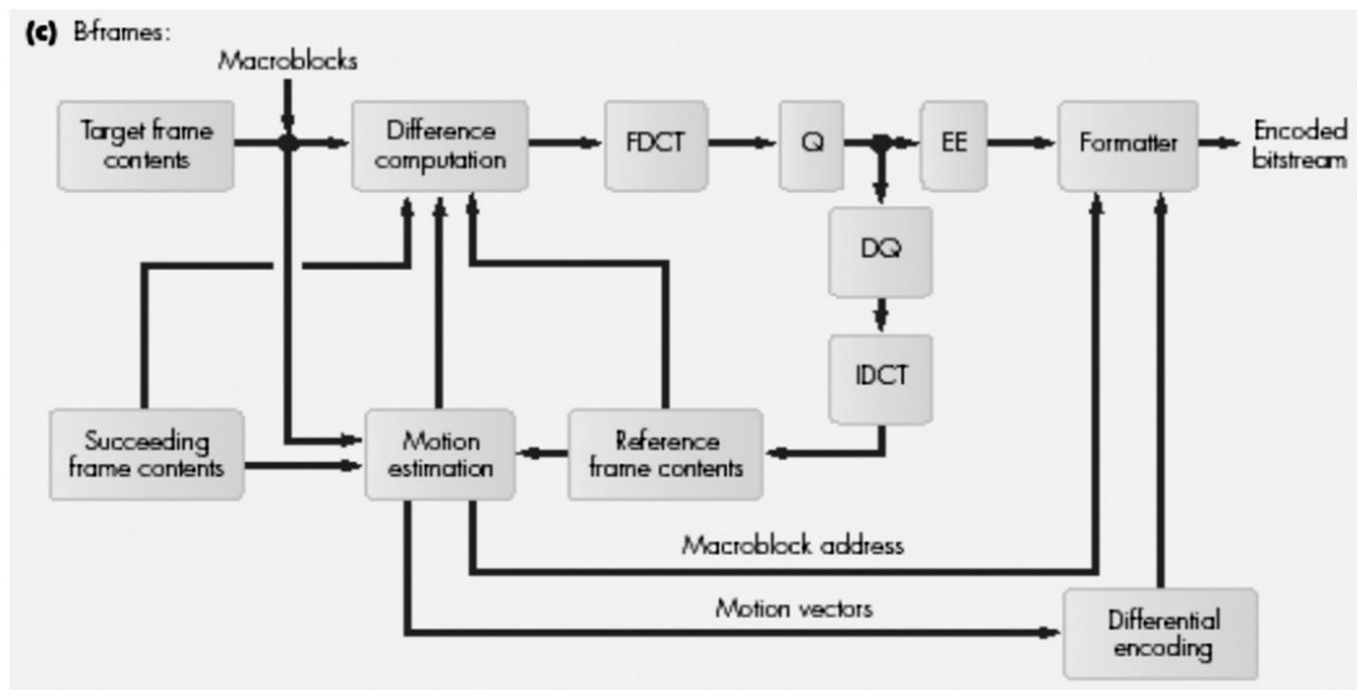
40

MPEG-1: I-Frame Encoding

I帧编码

(a) I-frames:

Macroblocks

Target frame contents → Forward DCT (FDCT) → Quantization (Q) → Entropy encoding (EE) → Formatter → Encoded bitstream

MPEG-1: P-Frame Encoding Flowchart

p帧编码流程图

宏模块
Macroblocks

不同的计算
量化
格式化程序

Target frame contents — Difference computation — FDCT — Q — EE — Formatter — Encoded bitstream

目标帧内容
后
前向离散余弦变换
熵编码
编码比特流

DQ
反量化

IDCT
逆离散余弦变换

后 - 前
Motion estimation
运动估计

Reference frame contents
参考帧内容
前

Macroblock address
宏模块地址

Motion vector
运动向量

Differential encoding
差分编码

Motion estimation

DQ = dequantizer   IDCT = inverse DCT

MPEG-1: B-Frame Encoding Flowchart

(c) B-frames:

MPEG-2: SNR Scalability 2:13:41

步长要小得多，因为你希望它覆盖 更多细节

SNR enhancement encoder

差异

基础编码器

Base encoder

Current frame

需要编码的帧
变为很多宏模块

预测误差

coarse quantization
步长更长

DCT

Q

Q^{-1}

VLC

Q

Q^{-1}

Bits_enhance

Bits_base

IDCT

重建你的参考系

基于运动补偿的预测

MC-based prediction

Frame memory

帧存储：以前的参考帧

Prediction
预测

最佳匹配块

Motion vectors

Motion estimation

运动估计

(a) Encoder
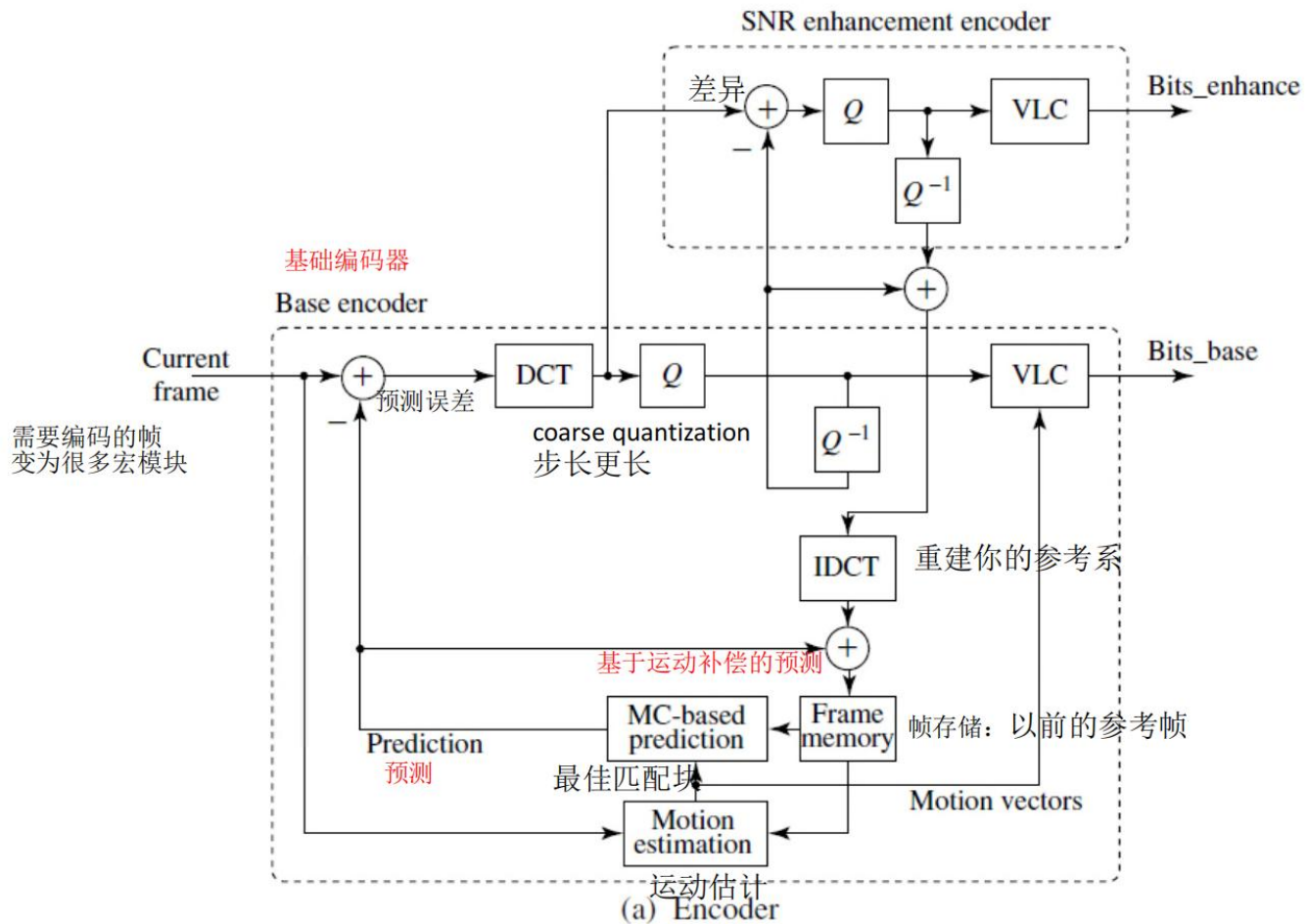
MPEG-2: Spatial Scalability

2:13:41

空间增强层编码器

Current frame

视频原始分辨率

Spatial enhancement layer encoder → Bits_enhance

Spatial decimator

空间抽取滤波器

低通滤波和采样

Spatial interpolator

空间内插程序

Spatial base layer encoder → Bits_base

基础层比特流

空间基础层编码器

(a)
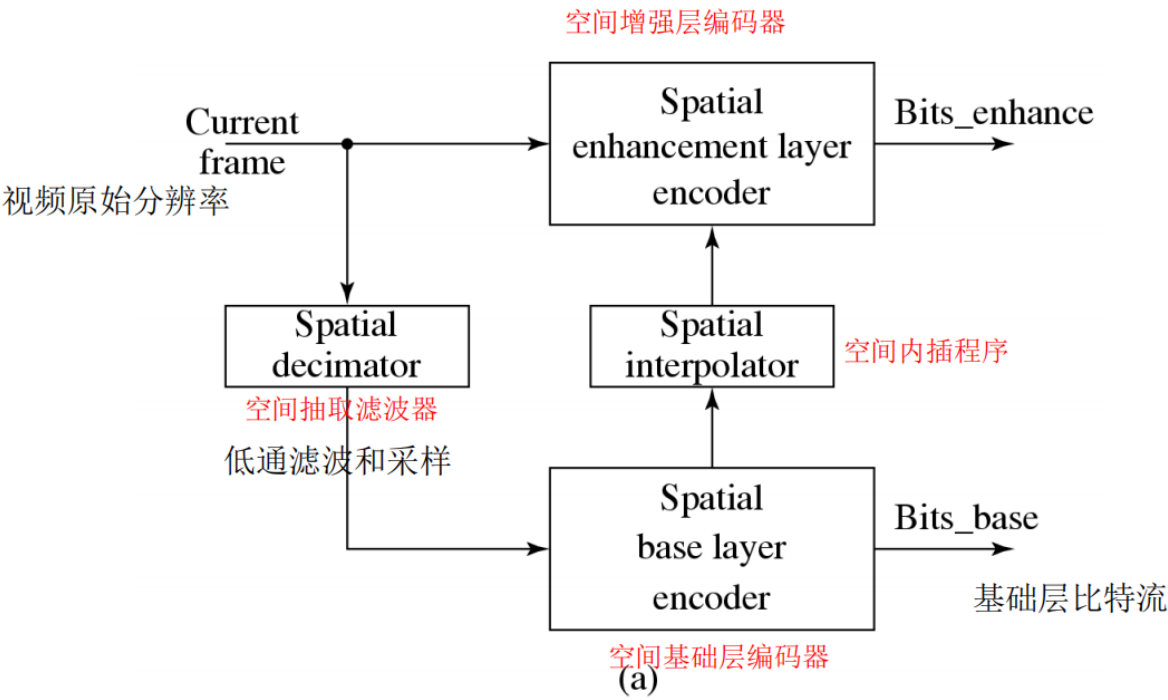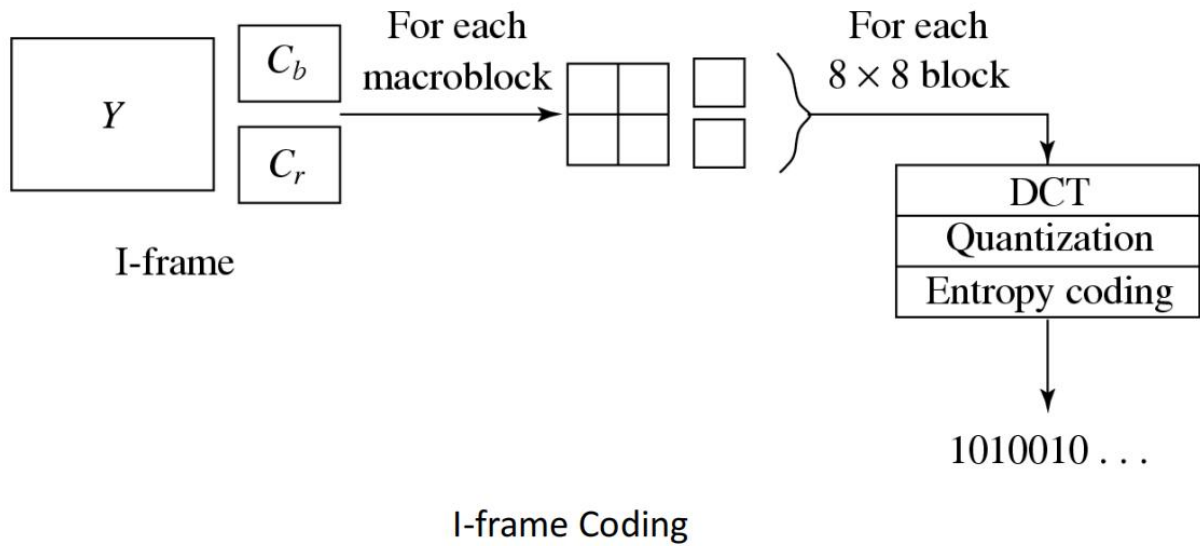
MPEG-2: Temporal Scalability
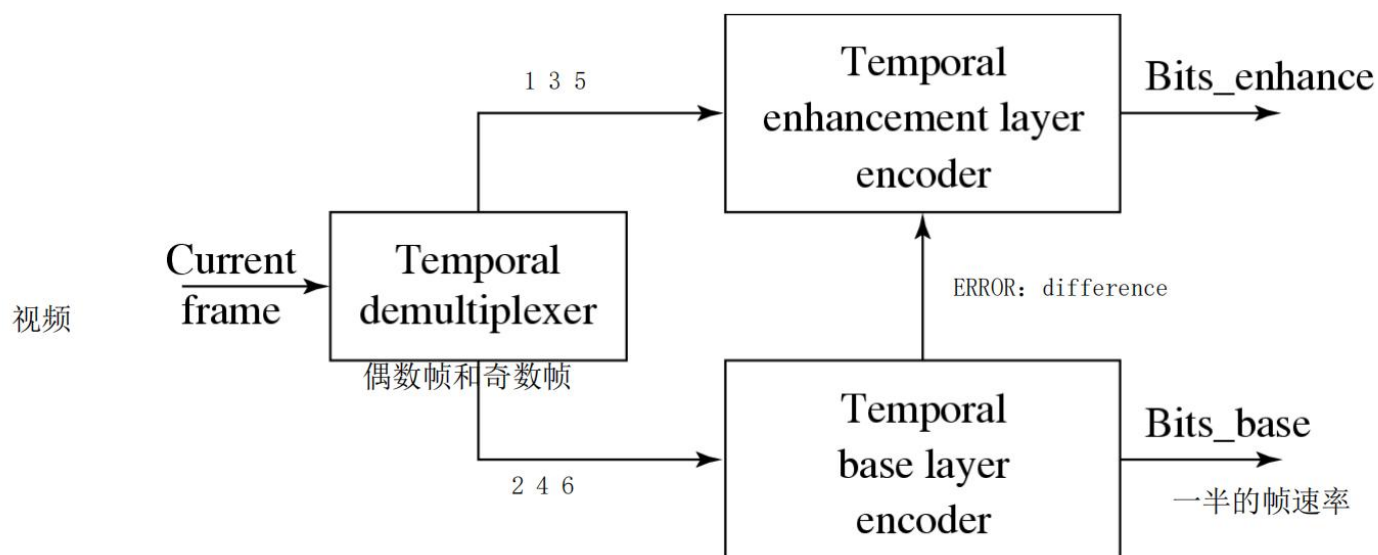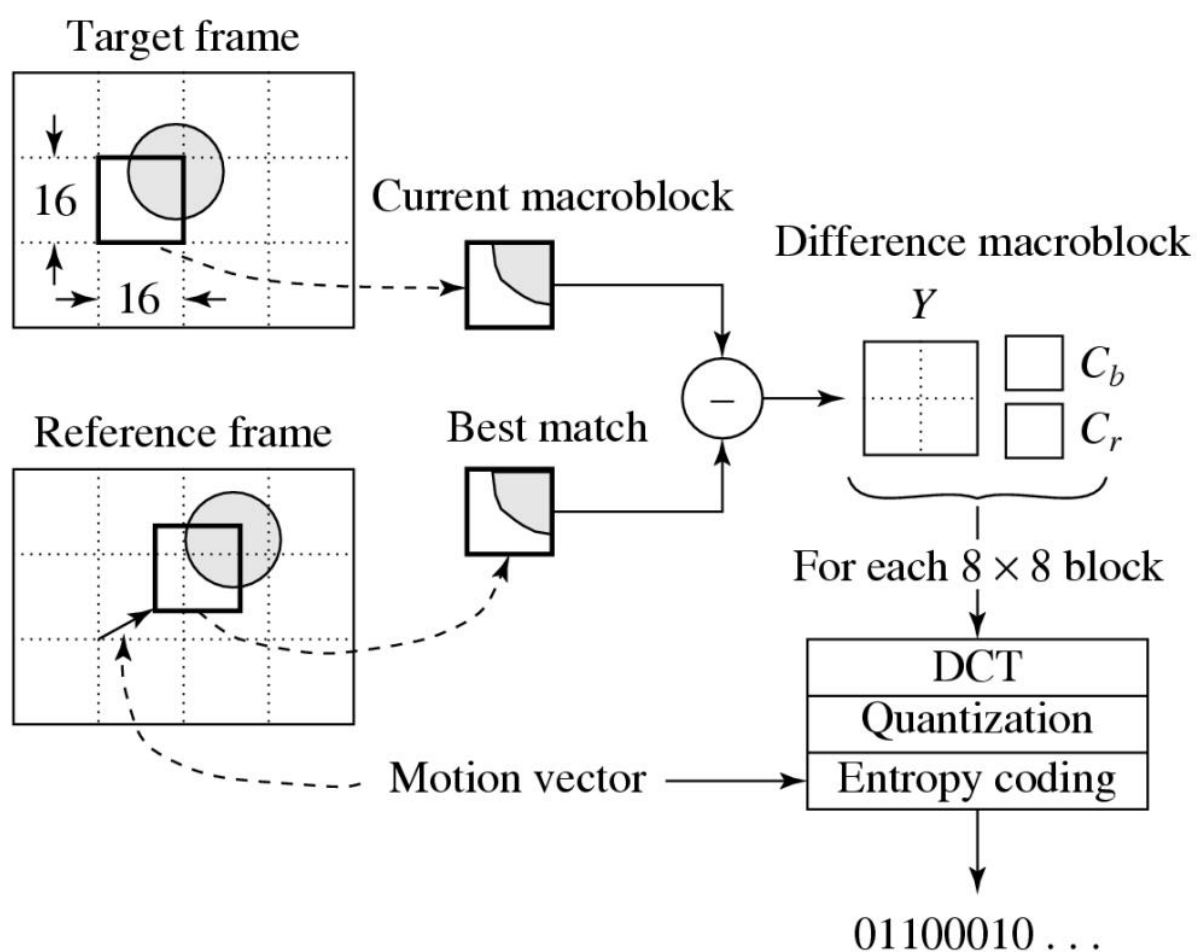

(a) Block Diagram

H.261: I-frame Coding

宏块包含4Y，1Cb和1Cr的8 x 8像素块，用于4:2:0色度子采样

- A macroblock contains 4Y, 1Cb, and 1Cr of 8 x 8 pixel blocks for 4:2:0 chroma subsampling.
- Macroblocks are of size 16 x 16 pixels for the Y frame, and 8 x 8 for Cb and Cr frames. 对于Y帧，macroblock的大小为16 x 16像素，对于badcrres为8 x 8像素。
- For each 8 x 8 block, a DCT transform is applied, the DCT coefficients then go through quantization, zigzag scanning, and entropy coding.

对于每个8 x 8块，应用DCT变换，DCT系数然后经过量化，之字形扫描和熵编码。

I-frame Coding

H.261: P-frame Coding

H.261 P-frame Coding Based on Motion Compensation

H.264: Encoder

编码器控件：控制生成的视频的比特率

输入视频帧

Input Video Frame Divided into 16×16 macroblocks

整数变换 Integer Transform

预测误差

Coder Control

Control Data

Scaling & Quantization
缩放与量化

量化变换系数

熵编码来生成比特流

Entropy Coding
熵编码

Output Code

编码的比特流 或压缩的视频

Scaling & Inv. Quant. 逆量化

Inv. Integer Transform 逆整数变换

预测误差的重建误差

加上前一帧

＋

得到下一帧

Intra−frame Prediction

Deblocking Filter 去块过滤器

Motion Compensation

预测误差

之前的参考帧

之前的参考系中最好的匹配块

Motion Estimation
运动估计

Motion Vectors 运动向量

120

ResNet

消失梯度问题

继续向普通 CNN 添加许多不同的层，实际上性能在提高一段时间后就会开始下降

如果你将梯度堆叠在许多不同的层上，是的，你的早期层无法有效更新

残差块

$\mathcal{F}(\mathbf{x})$

$\mathbf{x}$

weight layer

relu

weight layer

$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
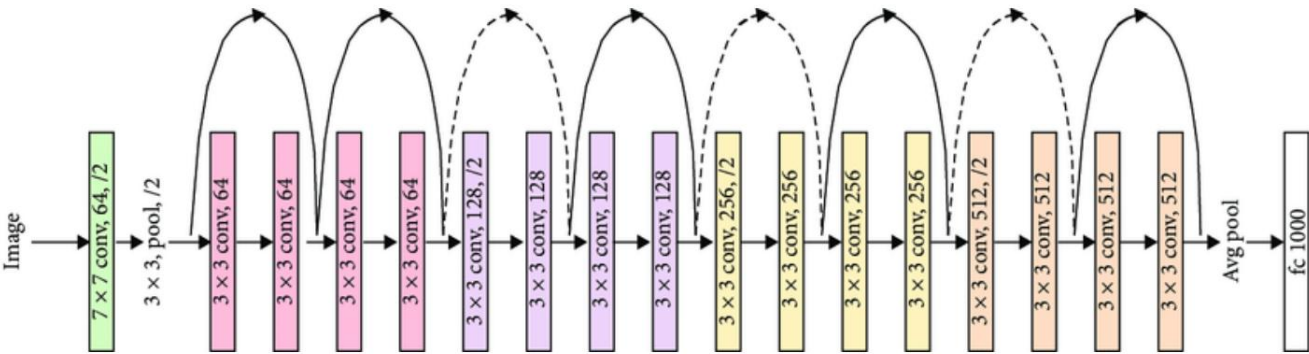
relu

$\mathbf{x}$

identity

高速公路或跳过连接
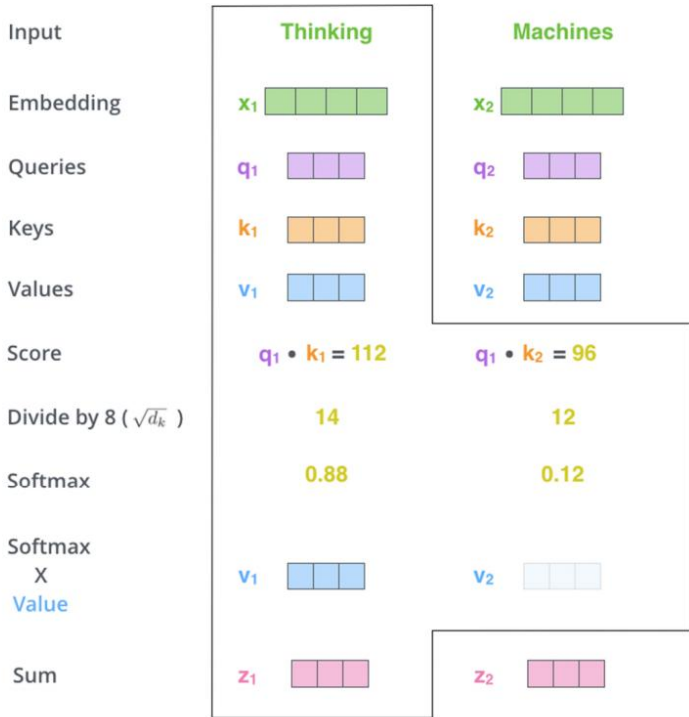
高速公路连接来更有效地反向传播这个梯度

仍然可以有效地反向传播

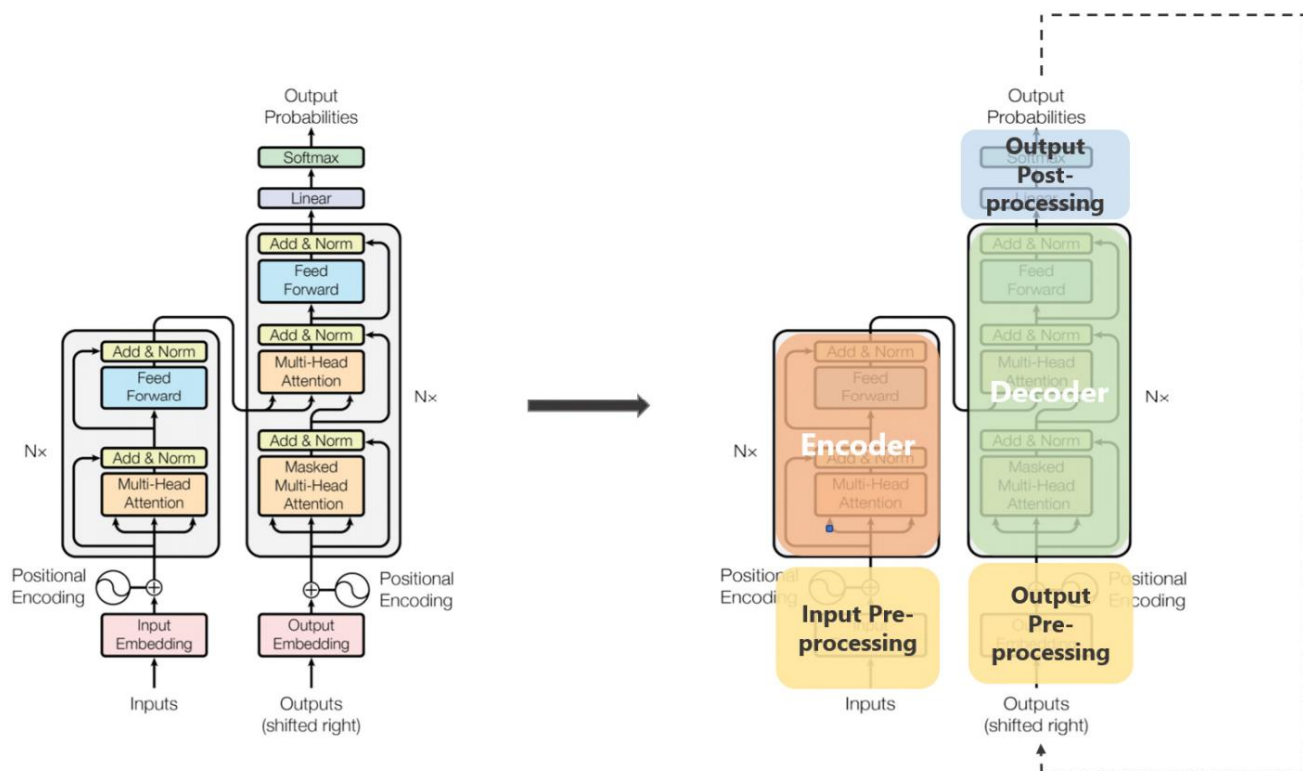# Scaled Dot-Product Attention 缩放的点积注意力

- Step 1: compute the correlation (dot product) between the query (q) and key (k) vectors.

- Step 2: correlation values from Step 1 are scaled and normalized using Softmax function.

- Step 3: multiplied output from Step 2 by corresponding value (v) vectors and sum them up.

步骤1:计算查询(q)和键(k)向量之间的相关性(点积)。
步骤2:使用Softmax函数对步骤1的相关值进行缩放和归一化。
步骤3:将步骤2的输出与相应的值向量(v)相乘并求和。

	Thinking	Machines
Input		
Embedding	x_1	x_2
Queries	q_1	q_2
Keys	k_1	k_2
Values	v_1	v_2
Score	$q_1 \cdot k_1 = 112$	$q_1 \cdot k_2 = 96$
Divide by 8 ($\sqrt{d_k}$)	14	12
Softmax	0.88	0.12
Softmax X Value	v_1	v_2
Sum	z_1	z_2

$$A(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

$$softmax\left(\frac{Q \times K^T}{\sqrt{d_k}}\right)V$$

$$= Z$$

125

Transformer Architecture (1) 变压器结构

Source: Ria Kulshrestha, Transformers. https://towardsdatascience.com/transformers-89034557de14