

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2025 年 6 月 26 日 (26.06.2025)



(10) 国际公布号
WO 2025/129752 A1

(51) 国际专利分类号:
G06T 3/40 (2024.01) **G06N 3/0455** (2023.01)

(21) 国际申请号: PCT/CN2023/143017

(22) 国际申请日: 2023 年 12 月 29 日 (29.12.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202311745925.4 2023年12月19日 (19.12.2023) CN

(71) 申请人: 华侨大学 (**HUAQIAO UNIVERSITY**) [CN/CN]; 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。厦门松霖科技股份有限公司 (**XIAMEN SOLEX HIGH-TECH INDUSTRIES CO., LTD.**) [CN/CN]; 中国福建省厦门市海沧区阳光西路 298 号 361028 (CN)。

(72) 发明人: 曾焕强 (**ZENG, Huanqiang**); 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。朱显丞 (**ZHU, Xiancheng**); 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。朱建清 (**ZHU, Jianqing**); 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。陈婧 (**CHEN, Jing**); 中国福建省泉州市丰

泽区城华北路 269 号 362011 (CN)。施一帆 (**SHI, Yifan**); 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。林琦 (**LIN, Qi**); 中国福建省泉州市丰泽区城华北路 269 号 362011 (CN)。陈斌 (**CHEN, Bin**); 中国福建省厦门市海沧区阳光西路 298 号 361028 (CN)。陈志达 (**CHEN, Zhida**); 中国福建省厦门市海沧区阳光西路 298 号 361028 (CN)。卢伟 (**LU, Wei**); 中国福建省厦门市海沧区阳光西路 298 号 361028 (CN)。夏至贤 (**HSIA, Chih-Hsien**); 中国台湾宜兰县宜兰市神农路一段 1 号 (CN)。

(74) 代理人: 厦门市首创君合专利事务所有限公司等 (**SHOUCHUANG JUNHE PATENT AGENT CO., LTD. XIAMEN et al.**); 中国福建省厦门市思明区软件园二期望海路 23 号之三 606 室 361012 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** SINGLE-FRAME IMAGE SUPER-RESOLUTION METHOD AND APPARATUS BASED ON HYBRID FEATURE INTERACTION TRANSFORMER

(54) 发明名称: 基于混合特征交互Transformer的单帧图像超分辨率方法及装置

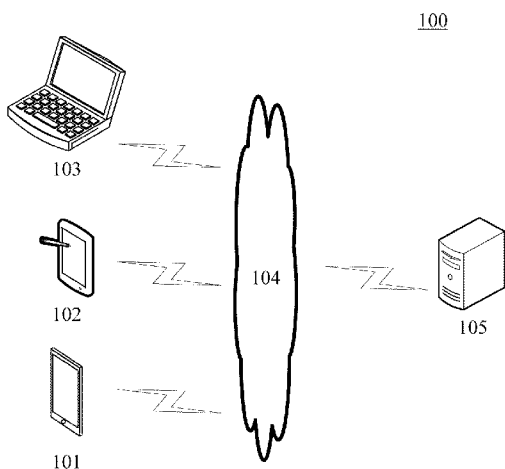


图 1

(57) **Abstract:** The present invention relates to the field of image processing. Disclosed are a single-frame image super-resolution (SR) method and apparatus based on a hybrid feature interaction transformer. The method comprises: acquiring a low-resolution image to be reconstructed; constructing a single-frame image SR model based on a hybrid feature interaction transformer and training same, so as to obtain a trained single-frame image SR model, wherein the single-frame image SR model comprises a shallow feature extraction unit, a deep feature extraction unit and an up-sampling reconstruction unit which are sequentially connected, and the deep feature extraction unit comprises P hybrid feature interaction transformer modules which are sequentially connected; and inputting the low-resolution image into the trained single-frame image SR model, extracting shallow features by means of the shallow feature extraction unit, inputting the shallow features into the deep feature extraction unit for extraction, so as to obtain deep features, and inputting the deep features into the up-sampling reconstruction unit for reconstruction to obtain a reconstructed high-resolution image. The problem of the reconstruction performance being affected due to an SR method for a transformer ignoring latent correlations between features from different dimensions is solved.

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚
(AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,
BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明公开了一种基于混合特征交互Transformer的单帧图像超分辨率方法及装置, 涉及图像处理领域, 包括: 获取待重建的低分辨率图像; 构建基于混合特征交互Transformer的单帧图像超分辨率模型并训练, 得到经训练的单帧图像超分辨率模型, 单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元, 深层特征提取单元包括依次连接的P个混合特征交互Transformer模块; 将低分辨率图像输入经训练的单帧图像超分辨率模型, 通过浅层特征提取单元提取浅层特征, 将浅层特征输入深层特征提取单元提取得到深层特征, 将深层特征输入上采样重建单元, 重建得到高分辨率重建图像, 解决Transformer的SR方法忽略不同维度的特征之间的潜在关联性, 影响重建性能的问题。

基于混合特征交互 Transformer 的单帧图像超分辨率方法及装置

技术领域

本发明涉及图像处理领域，具体涉及一种基于混合特征交互 Transformer 的单帧图像超分辨率方法及装置。

背景技术

图像超分辨率(image Super-Resolution, SR)是计算机视觉和图像处理的关键任务之一，它旨在从已有的低分辨率(Low-Resolution, LR)图像中重建高质量的高分辨率(High-Resolution, HR)图像。近来，基于卷积神经网络(Convolutional Neural Network, CNN)的 SR 方法由于其强大的特征表示、端到端的可训练范式、卓越的性能，在图像 SR 领域占据主导地位。然而，由于卷积操作通过使用固定的滑动窗口来提取小邻域内的局部特征，基于 CNN 的 SR 方法的信息性像素有限。当前，Transformer 作为一种新颖的 CNN 代替方案，在多种低级视觉任务上取得良好的表现。

对于图像 SR, Liang 等人提出了一种基于 Swin Transformer 的 SR 模型，即 SwinIR。SwinIR 采用了一种分层设计，将相似性计算限制在局部窗口中，并利用移动窗口机制来增强跨窗口的信息交互。然而，SwinIR 由于使用基于窗口的自注意力，放弃了全局信息推理，Transformer 的性能受到限制。为了激活更多有助于图像 SR 的信息性像素，Chen 等人提出了 HAT，其中引入通道注意力以更好地聚合跨窗口信息。Wang 等人提出了 Omni-SR，能够同时建模空间和窗口维度之间的像素级信息交互。然而，现有的基于 Transformer 的 SR 方法一般通过串行或并行的操作捕捉空间、通道间的关系，却忽略了不同维度的特征之间的潜在关联性，从而限制了基于 Transformer 的 SR 方法的性能。

发明内容

针对上述提到的技术问题。本申请的实施例的目的在于提出了一种基于混合特征交互 Transformer 的单帧图像超分辨率方法及装置，克服现有 Transformer 方法忽略不同维度的特征之间的潜在关联性的问题，通过鼓励跨维度的特征交互，显著提升图像超分辨率方法的全局特征表达能力和细节重建能力。

第一方面，本发明提供了一种基于混合特征交互 Transformer 的单帧图像超分辨率方法，包括以下步骤：

获取待重建的低分辨率图像；

构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型，单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元，深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；

将待重建的低分辨率图像输入经训练的单帧图像超分辨率模型，通过浅层特征提取单元提取浅层特征，将浅层特征输入深层特征提取单元提取得到深层特征，将深层特征输入上采样重建单元，重建得到高分辨率重建图像。

作为优选，混合特征交互 Transformer 模块包括高效局部特征提取单元、第一层标准化层、混合特征交互自注意力单元、第二层标准化层和混合尺度前馈神经网络，混合特征交互 Transformer 模块的输入依次经过高效局部特征提取单元和第一层标准化层，第一层标准化层的输出与混合特征交互 Transformer 模块的输入相加，得到第一混合特征，第一混合特征依次经过混合特征交互自注意力单元和第二层标准化层，第二层标准化层的输出与第一混合特征相加，得到第二混合特征，第二混合特征输入混合尺度前馈神经网络，得到混合特征交互 Transformer 模块的输出。

作为优选，高效局部特征提取单元包括依次连接的第一位移卷积层、第一 GeLU 激活函数层、第二位移卷积层、SE 模块和第三位移卷积层，高效局部特征提取单元的计算过程如下：

$$H_{ELF}(\cdot) = F_{shift-conv} \left(F_{SE} \left(F_{shift-conv} \left(GeLU \left(F_{shift-conv}(\cdot) \right) \right) \right) \right)$$

；

其中， $H_{ELF}(\cdot)$ 表示高效局部特征提取单元的函数， $F_{shift-conv}(\cdot)$ 表示第一位移卷积层、第二位移卷积层或第三位移卷积层的位移卷积操作， $F_{SE}(\cdot)$ 表示 SE 模块的函数， $GeLU(\cdot)$ 表示 GeLU 激活函数。

作为优选，混合特征交互自注意力单元包括局部窗口自注意力分支、深度卷

积分支和双向特征交互单元,双向特征交互单元包括空间交互单元和通道交互单元,通道交互单元包括依次连接的全局平均池化层、第一卷积层、第一批标准化层,第二 GeLU 激活函数层、第二卷积层和第一 Sigmoid 激活函数层,空间交互单元包括依次连接的第三卷积层、第二批标准化层、第三 GeLU 激活函数层、第四卷积层和第二 Sigmoid 激活函数,局部窗口自注意力分支包括查询线性层、键线性层、值线性层和局部窗口自注意力模块,深度卷积分支包括卷积核大小为 3×3 的第一深度卷积层,第一深度卷积层输出的局部特征输入通道交互单元,得到通道级动态权重,通道级动态权重输入局部窗口自注意力分支中对值线性层输出的值特征图进行自适应校正;局部窗口自注意力模块输出的全局特征输入空间交互单元,得到空间级动态权重,空间级动态权重输入深度卷积分支中对局部特征进行自适应校正。

作为优选,混合特征交互自注意力单元的计算过程如下:

将输入混合特征交互自注意力单元的第一特征图 $X \in \mathbb{R}^{C \times H \times W}$ 输入第一深度卷积层,得到局部特征 $F_{local} \in \mathbb{R}^{C \times H \times W}$,其中, $\mathbb{R}$ 表示实数多维空间, C 、 H 、 W 分别表示第一特征图的通道数量、长度和宽度, $\mathbb{R}^{C \times H \times W}$ 表示形状为 $C \times H \times W$ 的三维数据,窗口尺寸为 S ,其表达式如下:

$$F_{local} = DwConv_{3 \times 3}(X);$$

其中, $DwConv_{3 \times 3}(\cdot)$ 表示第一深度卷积层的函数;

将局部特征 F_{local} 输入通道交互单元,获得通道级动态权重

$W_{ca} \in \mathbb{R}^{C \times H \times W}$,其表达式如下:

$$W_{ca} = CI(F_{local});$$

其中, $CI(\cdot)$ 表示通道交互单元的函数;

将第一特征图 X 分割为 N 个窗口尺寸为 $S \times S$ 的非重叠窗口 $X_{win} \in \mathbb{R}^{NS^2 \times C}$ ，其中 $N = H \times W / S^2$ ， $\mathbb{R}^{NS^2 \times C}$ 表示形状为 $NS^2 \times C$ 的二维数据，分别通过查询线性层、键线性层、值线性层将非重叠窗口 X_{win} 分别转换为查询特征图 $Q \in \mathbb{R}^{NS^2 \times C}$ 、键特征图 $K \in \mathbb{R}^{NS^2 \times C}$ 和值特征图 $V \in \mathbb{R}^{NS^2 \times C}$ ，其表达式如下：

$$Q, K, V = L_Q(X_{win}), L_K(X_{win}), L_V(X_{win});$$

其中， L_Q 、 L_K 、 L_V 分别表示查询线性层、键线性层、值线性层的函数；

将值特征图 V 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ 并与通道级动态权重 W^{ca} 相乘，以对值特征图 V 进行自适应校正，随后再将数据格式还

原为 $NS^2 \times C$ ，校正后的结果记作 V' ；

执行局部窗口自注意力模块的计算，以获得全局特征 $F_{global} \in \mathbb{R}^{NS^2 \times C}$ ，表达式如下：

$$F_{global} = \text{Softmax} \left(\frac{QK^T}{\sqrt{C}} V' \right);$$

其中， T 表示转置矩阵， Softmax 表示 Softmax 函数；

将全局特征 F_{global} 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ ，并

输入空间交互单元，以获得空间级动态权重 $W_{sa} \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$W_{sa} = SI(F_{global});$$

其中, $SI(\cdot)$ 表示空间交互单元的函数;

通过将空间级动态权重 W_{sa} 与全局特征 F_{local} 相乘, 以对全局特征

F_{local} 进行自适应校正, 其校正后的结果记作 F'_{local} ;

最后, 将全局特征 F_{local} 与 F'_{local} 相加, 以获得混合特征

$$F_{mix} \in \mathbb{R}^{C \times H \times W}.$$

作为优选, 混合尺度前馈神经网络包括第一分支、第二分支和第五卷积层, 第一分支包括依次连接的第二深度卷积层、第一 ReLU 激活函数层、第三深度卷积层和第二 ReLU 激活函数层, 第二分支包括依次连接的第四深度卷积层、第三 ReLU 激活函数层、第五深度卷积层和第四 ReLU 激活函数层, 其中第二深度卷积层和第五深度卷积层的卷积核大小为 7×7 , 第三深度卷积层和第四深度卷积层的卷积核大小为 5×5 , 具体计算过程如下:

沿着通道维度将输入混合尺度前馈神经网络的第二特征图 $Y' \in \mathbb{R}^{C \times H \times W}$

将 Y' 切分为均等的两份, 得到切分后的特征 $X_{slice}^1 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 和

$X_{slice}^2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$, $\mathbb{R}^{\frac{C}{2} \times H \times W}$ 表示形状为 $\frac{C}{2} \times H \times W$ 的三维数据,

将 X_{slice}^1 和 X_{slice}^2 分别输入第一分支和第二分支进行混合交叉特征提取, 分别输出第一交叉特征和第二交叉特征, 将第一交叉特征和第二交叉特征进行拼接后输入第五卷积层, 第五卷积层的输出与第二特征图相加, 得到混合尺度特征

$Y \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$X_1^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_{\text{slice}}^1\right)\right);$$

$$X_1^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_{\text{slice}}^2\right)\right);$$

$$X_2^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$X_2^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$Y = \text{Conv}_{1 \times 1}\left([X_2^{p_1}, X_2^{p_2}]\right) + X';$$

其中， $\text{ReLU}(\cdot)$ 表示 ReLU 激活函数， $\text{DwConv}_{5 \times 5}(\cdot)$ 和

$\text{DwConv}_{7 \times 7}(\cdot)$ 分别表示卷积核为 5×5 和 7×7 的深度卷积层的函数，

$\text{Conv}_{1 \times 1}(\cdot)$ 表示卷积核大小为 1×1 的第五卷积层的函数， $[\cdot]$ 表示拼接操作，

$X_1^{p_1}$ 、 $X_1^{p_2}$ 分别表示第一特征和第二特征， $X_2^{p_1}$ 、 $X_2^{p_2}$ 分别表示第一交叉特征和第二交叉特征。

作为优选，单帧图像超分辨率模型的具体结构和计算过程如下：

浅层特征提取单元采用第六卷积层，浅层特征提取单元的计算过程如下：

$$F_0 = \text{Conv}_{3 \times 3}\left(I_{LR}\right);$$

其中， F_0 表示浅层特征 $F_0 \in \mathbb{R}^{C \times H \times W}$ ， $\text{Conv}_{3 \times 3}(\cdot)$ 表示卷积

核为 3×3 的第六卷积层的函数, I_{LR} 表示低分辨率图像;

利用 P 个混合特征交互 Transformer 模块提取特征, 并利用长跳跃连接将 F_0 传递至网络末端, 与第 P 个混合特征交互 Transformer 模块的输出相加以进行残差学习, 其表达式如下:

$$F_i = MF^i(F_{i-1}), i \in [1, P];$$

$$F_{P0} = MF^P(\dots(MF^2(MF^1(F_0)))) + F_0;$$

其中, F_{i-1} 表示第 $i-1$ 个所述混合特征交互 Transformer 模块的输出, MF^P 表示第 P 个所述混合特征交互 Transformer 模块的函数, MF^1 表示第 1 个所述混合特征交互 Transformer 模块的函数, MF^2 表示第 2 个所述混合特征交互 Transformer

模块的函数, MF^i 表示第 i 个混合特征交互 Transformer 模块的函数, F_i 表示第 i 个混合特征交互 Transformer 模块的输出, F_{P0} 表示深层特征, $F_{P0} \in \mathbb{R}^{C \times H \times W}$;

上采样重建单元包括尺度因子为 $scale$ 的亚像素卷积层和卷积核为 3×3 的第七卷积层, 其表达式如下:

$$I_{SR} = Conv_{3 \times 3}(f_{up}(F_{P0}));$$

其中, $f_{up}(\cdot)$ 表示亚像素卷积层的函数, $Conv_{3 \times 3}(\cdot)$ 表示第七卷积层的

函数, I_{SR} 表示高分辨率重建图像, $I_{SR} \in \mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$,

$\mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$ 表示形状为 $3 \times (H \times scale) \times (W \times scale)$ 的三维数据。

第二方面, 本发明提供了一种基于混合特征交互 Transformer 的单帧图像超分辨率装置, 包括:

图像获取模块, 被配置为获取待重建的低分辨率图像;

模型构建模块，被配置为构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型，单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元，深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；

重建模块，被配置为将待重建的低分辨率图像输入经训练的单帧图像超分辨率模型，通过浅层特征提取单元提取浅层特征，将浅层特征输入深层特征提取单元提取得到深层特征，将深层特征输入上采样重建单元，重建得到高分辨率重建图像。

第三方面，本发明提供了一种电子设备，包括一个或多个处理器；存储装置，用于存储一个或多个程序，当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

相比于现有技术，本发明具有以下有益效果：

(1) 本发明提出的基于混合特征交互 Transformer 的单帧图像超分辨率方法中的混合特征交互自注意力单元采用双分支结构结合双向特征交互单元，双分支结构在标准的局部窗口自注意力单元的基础上，额外引入一个与局部窗口自注意力单元平行的深度卷积分支，能够增强 Transformer 的跨窗口特征交互能力，该双向特征交互单元能为双分支结构提供互补的线索，充分考虑不同类型特征之间的互补性，能够显著提升信息利用率和图像超分辨率性能。

(2) 本发明提出的基于混合特征交互 Transformer 的单帧图像超分辨率方法能够克服现有 Transformer 方法忽略不同维度的特征之间的潜在关联性的问题，通过鼓励跨维度的特征交互，显著提升图像超分辨率方法的全局特征表达能力和细节重建能力。

(3) 本发明提出的基于混合特征交互 Transformer 的单帧图像超分辨率方法与现有的单帧图像超分辨率方法相比，具有较低的参数量和 Flops 值，综合性能最优，能够利用更少的计算资源实现高性能的图像超分辨率重建。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简要介绍，显而易见地，下面描述中的附图仅仅是本发

明的一些实施例，对于本领域的普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请的一个实施例可以应用于其中的示例性装置架构图；

图 2 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的流程示意图；

图 3 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的高效局部特征提取单元的结构示意图；

图 4 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的混合特征交互自注意力单元的结构示意图；

图 5 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的混合尺度前馈神经网络的结构示意图；

图 6 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的混合特征交互 Transformer 模块的结构示意图；

图 7 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法的基于混合特征交互 Transformer 的单帧图像超分辨率模型的结构示意图；

图 8 为本申请的实施例的基于混合特征交互 Transformer 的单帧图像超分辨率装置的示意图；

图 9 是适于用来实现本申请实施例的电子设备的计算机装置的结构示意图。

具体实施方式

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进一步地详细描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。

图 1 示出了可以应用本申请实施例的基于混合特征交互 Transformer 的单帧图像超分辨率方法或基于混合特征交互 Transformer 的单帧图像超分辨率装置的示例性装置架构 100。

如图 1 所示, 装置架构 100 可以包括终端设备一 101、终端设备二 102、终端设备三 103, 网络 104 和服务器 105。网络 104 用以在终端设备一 101、终端设备二 102、终端设备三 103 和服务器 105 之间提供通信链路的介质。网络 104 可以包括各种连接类型, 例如有线、无线通信链路或者光纤电缆等等。

需要说明的是, 本申请实施例所提供的基于混合特征交互 Transformer 的单帧图像超分辨率方法可以由服务器 105 执行, 也可以由终端设备一 101、终端设备二 102、终端设备三 103 执行, 相应地, 基于混合特征交互 Transformer 的单帧图像超分辨率装置可以设置于服务器 105 中, 也可以设置于终端设备一 101、终端设备二 102、终端设备三 103 中。

应该理解, 图 1 中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要, 可以具有任意数目的终端设备、网络和服务器的。在所处理的数据不需要从远程获取的情况下, 上述装置架构可以不包括网络, 而只需服务器或终端设备。

图 2 示出了本申请的实施例提供的一种基于混合特征交互 Transformer 的单帧图像超分辨率方法, 包括以下步骤:

S1, 获取待重建的低分辨率图像。

具体的, 收集待重建的低分辨率图像, 该低分辨率图像为单帧图像。

S2, 构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练, 得到经训练的单帧图像超分辨率模型, 单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元, 深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块。

在具体的实施例中, 混合特征交互 Transformer 模块包括高效局部特征提取单元、第一层标准化层、混合特征交互自注意力单元、第二层标准化层和混合尺度前馈神经网络, 混合特征交互 Transformer 模块的输入依次经过高效局部特征提取单元和第一层标准化层, 第一层标准化层的输出与混合特征交互 Transformer 模块的输入相加, 得到第一混合特征, 第一混合特征依次经过混合特征交互自注意力单元和第二层标准化层, 第二层标准化层的输出与第一混合特征相加, 得到第二混合特征, 第二混合特征输入混合尺度前馈神经网络, 得到混合特征交互

Transformer 模块的输出。

在具体的实施例中，高效局部特征提取单元包括依次连接的第一位移卷积层、第一 GeLU 激活函数层、第二位移卷积层、SE 模块和第三位移卷积层，高效局部特征提取单元的计算过程如下：

$$H_{ELF}(\cdot) = F_{shift-conv} \left(F_{SE} \left(F_{shift-conv} \left(GeLU \left(F_{shift-conv}(\cdot) \right) \right) \right) \right)$$

；

其中， $H_{ELF}(\cdot)$ 表示高效局部特征提取单元的函数， $F_{shift-conv}(\cdot)$ 表示第一位移卷积层、第二位移卷积层或第三位移卷积层的位移卷积操作， $F_{SE}(\cdot)$ 表示 SE 模块的函数， $GeLU(\cdot)$ 表示 GeLU 激活函数。

在具体的实施例中，混合特征交互自注意力单元包括局部窗口自注意力分支、深度卷积分支和双向特征交互单元，双向特征交互单元包括空间交互单元和通道交互单元，通道交互单元包括依次连接的全局平均池化层、第一卷积层、第一批标准化层，第二 GeLU 激活函数层、第二卷积层和第一 Sigmoid 激活函数层，空间交互单元包括依次连接的第三卷积层、第二批标准化层、第三 GeLU 激活函数层、第四卷积层和第二 Sigmoid 激活函数，局部窗口自注意力分支包括查询线性层、键线性层、值线性层和局部窗口自注意力模块，深度卷积分支包括卷积核大小为 3×3 的第一深度卷积层，第一深度卷积层输出的局部特征输入通道交互单元，得到通道级动态权重，通道级动态权重输入局部窗口自注意力分支中对值线性层输出的值特征图进行自适应校正；局部窗口自注意力模块输出的全局特征输入空间交互单元，得到空间级动态权重，空间级动态权重输入深度卷积分支中对局部特征进行自适应校正。

在具体的实施例中，混合特征交互自注意力单元的计算过程如下：

将输入混合特征交互自注意力单元的第一特征图 $X \in \mathbb{R}^{C \times H \times W}$ 输入第一深度卷积层，得到局部特征 $F_{local} \in \mathbb{R}^{C \times H \times W}$ ，其中， $\mathbb{R}$ 表示实数多维空间， C, H, W 分别表示第一特征图的通道数量、长度和宽度，窗口尺寸为 S ， $\mathbb{R}^{C \times H \times W}$

表示形状为 $C \times H \times W$ 的三维数据，其表达式如下：

$$F_{local} = DwConv_{3 \times 3}(X);$$

其中， $DwConv_{3 \times 3}(\cdot)$ 表示第一深度卷积层的函数；

将局部特征 F_{local} 输入通道交互单元，获得通道级动态权重

$W_{ca} \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$W_{ca} = CI(F_{local});$$

其中， $CI(\cdot)$ 表示通道交互单元的函数；

将第一特征图 X 分割为 N 个窗口尺寸为 $S \times S$ 的非重叠窗口

$X_{win} \in \mathbb{R}^{NS^2 \times C}$ ，其中 $N = H \times W / S^2$ ， $\mathbb{R}^{NS^2 \times C}$ 表示形状为

$NS^2 \times C$ 的二维数据，分别通过查询线性层、键线性层、值线性层将非重

叠窗口 X_{win} 分别转换为查询特征图 $Q \in \mathbb{R}^{NS^2 \times C}$ 、键特征图

$K \in \mathbb{R}^{NS^2 \times C}$ 和值特征图 $V \in \mathbb{R}^{NS^2 \times C}$ ，其表达式如下：

$$Q, K, V = L_Q(X_{win}), L_K(X_{win}), L_V(X_{win});$$

其中， L_Q 、 L_K 、 L_V 分别表示查询线性层、键线性层、值线性层的函数；

将值特征图 V 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ 并与通道

级动态权重 W_{ca} 相乘，以对值特征图 V 进行自适应校正，随后再将数据格式还

原为 $NS^2 \times C$ ，校正后的结果记作 V' ；

执行局部窗口自注意力模块的计算，以获得全局特征

$F_{global} \in \mathbb{R}^{NS^2 \times C}$ ，表达式如下：

$$F_{global} = \text{Softmax} \left(\frac{QK^T}{\sqrt{C}} V' \right);$$

其中，T 表示转置矩阵，Softmax 表示 Softmax 函数；

将全局特征 F_{global} 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ ，并

输入空间交互单元，以获得空间级动态权重 $W_{sa} \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$W_{sa} = SI(F_{global});$$

其中， $SI(\cdot)$ 表示空间交互单元的函数；

通过将空间级动态权重 W_{sa} 与全局特征 F_{local} 相乘，以对全局特征

F_{local} 进行自适应校正，其校正后的结果记作 F'_{local} ；

最后，将全局特征 F_{local} 与 F'_{local} 相加，以获得混合特征

$$F_{mix} \in \mathbb{R}^{C \times H \times W}。$$

在具体的实施例中，混合尺度前馈神经网络包括第一分支、第二分支和第五卷积层，第一分支包括依次连接的第二深度卷积层、第一 ReLU 激活函数层、第三深度卷积层和第二 ReLU 激活函数层，第二分支包括依次连接的第四深度卷积

层、第三 ReLU 激活函数层、第五深度卷积层和第四 ReLU 激活函数层，其中第二深度卷积层和第五深度卷积层的卷积核大小为 7×7 ，第三深度卷积层和第四深度卷积层的卷积核大小为 5×5 ，具体计算过程如下：

沿着通道维度将输入混合尺度前馈神经网络的第二特征图 $X' \in \mathbb{R}^{C \times H \times W}$

将 X' 切分为均等的两份，得到切分后的特征 $X_{slice}^1 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 和

$X_{slice}^2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ ， $\mathbb{R}^{\frac{C}{2} \times H \times W}$ 表示形状为 $\frac{C}{2} \times H \times W$ 的三维数据，

将 X_{slice}^1 和 X_{slice}^2 分别输入第一分支和第二分支进行混合交叉特征提取，分别输出第一交叉特征和第二交叉特征，将第一交叉特征和第二交叉特征进行拼接后输入第五卷积层，第五卷积层的输出与第二特征图相加，得到混合尺度特征

$Y \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$X_1^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_{slice}^1\right)\right);$$

$$X_1^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_{slice}^2\right)\right);$$

$$X_2^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$X_2^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$Y = \text{Conv}_{1 \times 1}\left([X_2^{p_1}, X_2^{p_2}]\right) + X';$$

其中， $\text{ReLU}(\cdot)$ 表示 ReLU 激活函数， $\text{DwConv}_{5 \times 5}(\cdot)$ 和

$DwConv_{7 \times 7}(\cdot)$ 分别表示卷积核为 5×5 和 7×7 的深度卷积层的函数，

$Conv_{1 \times 1}(\cdot)$ 表示卷积核大小为 1×1 的第五卷积层的函数， $\begin{bmatrix} \cdot \end{bmatrix}$ 表示拼接操作， $Y_1^{p_1}$ 、 $Y_1^{p_2}$ 分别表示第一特征和第二特征， $Y_2^{p_1}$ 、 $Y_2^{p_2}$ 分别表示第一交叉特征和第二交叉特征。

具体的，参考图 3，可先构建高效局部特征提取单元，该高效局部特征提取单元依次由第一位移卷积层、第一 GeLU 激活函数层、第二位移卷积层、SE 模块和第三位移卷积层级联构成，其中，SE 模块为压缩-激励模块

(Squeeze-Excitation Module)。参考图 4，再构建混合特征交互自注意力单元，混合特征交互自注意力单元在标准的局部窗口自注意力单元的基础上，通过增加两个关键设计而构建：（1）双分支结构，包括局部窗口自注意力分支和深度卷积分支；（2）双向特征交互单元。具体地，通过设计一个简洁的双分支结构，为标准的局部窗口自注意力单元引入一个与其平行的深度卷积层，以增强跨窗口的特征交互。双向特征交互单元包括空间交互单元和通道交互单元，深度卷积分支的信息首先通过空间交互单元流入局部窗口自注意力分支；然后，局部窗口自注意力分支的信息通过空间交互单元流入深度卷积分支。因此，本申请的实施例所提出的双向特征交互单元能为双分支结构提供互补的线索，以增强信息利用率。具体的，该通道交互单元由全局平均池化层、卷积核大小为 3×3 的第一卷积层、第一批标准化层，第二 GeLU 激活函数层、卷积核大小为 3×3 的第二卷积层和第一 Sigmoid 激活函数层级联构成。该空间交互单元由卷积核大小为 3×3 的第三卷积层、第二批标准化层、第三 GeLU 激活函数层、卷积核大小为 3×3 的第四卷积层和第二 Sigmoid 激活函数级联构成。而后，参考图 5，再构建混合尺度前馈神经网络包括两条多尺度深度卷积分支，该两条多尺度深度卷积分支通过交替使用卷积核大小为 5×5 的深度卷积层、卷积核大小为 7×7 的深度卷积层实现混合特征提取，每个深度卷积层后面均连接一个 ReLU 激活函数层，最后利用卷积核大小为 1×1 的第五卷积层融合两条分支的输出，得到混合尺度特征

$$Y \in \mathbb{R}^{C \times H \times W}.$$

进一步的，参考图 6，整合高效局部特征提取单元、混合特征交互自注意力单元和混合尺度前馈神经网络构建混合特征交互 Transformer 模块，该混合特征交互 Transformer 模块通过依次由高效局部特征提取单元、第一层标准化层、混合特征交互自注意力单元、第二层标准化层、混合尺度前馈神经网络级联构成。

最后，参考图 7，构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型。

S3，将待重建的低分辨率图像输入经训练的单帧图像超分辨率模型，通过浅层特征提取单元提取浅层特征，将浅层特征输入深层特征提取单元提取得到深层特征，将深层特征输入上采样重建单元，重建得到高分辨率重建图像。

在具体的实施例中，单帧图像超分辨率模型的具体结构和计算过程如下：

浅层特征提取单元采用第六卷积层，浅层特征提取单元的计算过程如下：

$$F_0 = \text{Conv}_{3 \times 3}(I_{LR}) ;$$

其中， F_0 表示浅层特征 $F_0 \in \mathbb{R}^{C \times H \times W}$ ， $\text{Conv}_{3 \times 3}(\cdot)$ 表示卷积

核为 3×3 的第六卷积层的函数， I_{LR} 表示低分辨率图像；

利用 P 个混合特征交互 Transformer 模块提取特征，并利用长跳跃连接将 F_0 传递至网络末端，与第 P 个混合特征交互 Transformer 模块的输出相加以进行残差学习，其表达式如下：

$$F_i = MF^i(F_{i-1}), i \in [1, P] ;$$

$$F_{P0} = MF^P(\dots(MF^2(MF^1(F_0)))) + F_0 ;$$

其中， F_{i-1} 表示第 $i-1$ 个所述混合特征交互 Transformer 模块的输出， MF^P 表示第 P 个所述混合特征交互 Transformer 模块的函数， MF^1 表示第 1 个所述混合特征交互 Transformer 模块的函数， MF^2 表示第 2 个所述混合特征交互 Transformer

模块的函数， MF^i 表示第 i 个混合特征交互 Transformer 模块的函数， F_i 表示第

i 个混合特征交互 Transformer 模块的输出, F_{P0} 表示深层特征, $F_{P0} \in \mathbb{R}^{C \times H \times W}$;

上采样重建单元包括尺度因子为 $scale$ 的亚像素卷积层和卷积核为 3×3 的第七卷积层, 其表达式如下:

$$I_{SR} = Conv_{3 \times 3} \left(f_{up} (F_{P0}) \right);$$

其中, $f_{up}(\cdot)$ 表示亚像素卷积层的函数, $Conv_{3 \times 3}(\cdot)$ 表示第七卷积层

的函数, I_{SR} 表示高分辨率重建图像, $I_{SR} \in \mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$,

$\mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$ 表示形状为 $3 \times (H \times scale) \times (W \times scale)$ 的三维数据。

具体的, 使用经训练的单帧图像超分辨率模块对待重建的低分辨率图像进行重建, 得到重建结果。该经训练的单帧图像超分辨率模块由三个部分组成: 浅层特征提取单元、深层特征提取单元、上采样重建单元。针对给定的待重建的低分

分辨率图像 $I_{LR} \in \mathbb{R}^{3 \times H \times W}$, 缩放因子 $scale$ 作为输入, 其中, $scale$ 的值为需求的放大倍数, 例如 $scale$ 为 2、3、4 或 8。

将本申请的实施例提出的一种基于混合特征交互 Transformer 的单帧图像超分辨率方法与目前最先进的单帧图像超分辨率方法进行比较。在本对比实验中, 使用 DIV2K 作为训练集, Set5、Set14、BSD100 和 Urban100 作为测试集, 目标缩放因子为 2。利用定量指标 PSNR 和 SSIM 评价重建图像质量, 越高的 PSNR 和 SSIM 数值对应越高的 SR 性能。利用定量指标参数量(Params)和 FLOPs 衡量模型规模和执行速度, 参数量越低, 表示模型规模越小, FLOPs 值越低, 表示模型执行速度越快。为了满足真实应用场景的需求, 设计具有低参数量、低 FLOPs 值却能生成具有高 PSNR 和 SSIM 的重建图像的图像超分辨率方法是图像超分辨率领域的重要目标。如表 1 所示, 与其他方法相比, 本申请的实施例所提出的方法以最低的参数量、最低的 FLOPs 值在四个测试集获得最高的 PSNR 和 SSIM。因

此,表 1 充分说明了本申请的实施例所提出的一种基于混合特征交互 Transformer 的单帧图像超分辨率方法与其他方法相比,展示了最优的综合性能。

表 1

方法		参数量	Flops	Set5	Set14	BSD100	Urban100
				PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
×2 SR	CARN	1592K	222.8G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256
	IMDN	694K	158.8G	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283
	MADNet	878K	187G	37.85/0.9600	33.38/0.9161	32.04/0.8979	31.62/0.9233
	LatticeNet	756K	170G	38.15/0.9610	33.78/0.9193	32.25/0.9005	32.43/0.9302
	LAPAR-A	548K	171G	38.01/0.9605	33.62/0.9183	32.19/0.8999	32.10/0.9283
	SwinIR	910K	244G	38.14/0.9611	33.86/0.9206	32.31/0.9012	32.76/0.9340
	ESRT	677K	234G	38.03/0.9600	33.75/0.9184	32.25/0.9001	32.58/0.9318
	ELAN	582K	168G	38.17/0.9611	33.94/0.9207	32.30/0.9012	32.76/0.9340
	SPIN	497K	-	38.20/0.9615	33.90/0.9215	32.31/0.9015	32.79/0.9340
	MSRA-SR	769K	196G	38.23/0.9614	34.01 / 0.9211	32.33 /0.9017	32.98/0.9358
	NGswin	998K	140G	38.05/0.9610	33.79/0.9199	32.27/0.9008	32.53/0.9324
	OmniSR	771K	153G	38.22/0.9613	33.98/0.9210	32.36/0.9020	33.05/0.9363
	SRFormer	853K	236G	38.23/0.9613	33.94/0.9209	32.36/0.9019	32.91/0.9353
	本申请	398K	78G	38.28/0.9627	34.98/0.9227	32.41/ 0.9023	33.17/0.9385

以上步骤 S1-S3 并不仅仅代表步骤之间的顺序,而是步骤符号表示。

进一步参考图 8,作为对上述各图所示方法的实现,本申请提供了一种基于混合特征交互 Transformer 的单帧图像超分辨率装置的一个实施例,该装置实施例与图 2 所示的方法实施例相对应,该装置具体可以应用于各种电子设

备中。

本申请实施例提供了一种基于混合特征交互 Transformer 的单帧图像超分辨率装置，包括：

图像获取模块 1，被配置为获取待重建的低分辨率图像；

模型构建模块 2，被配置为构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型，单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元，深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；

重建模块 3，被配置为将待重建的低分辨率图像输入经训练的单帧图像超分辨率模型，通过浅层特征提取单元提取浅层特征，将浅层特征输入深层特征提取单元提取得到深层特征，将深层特征输入上采样重建单元，重建得到高分辨率重建图像。

下面参考图 9，其示出了适于用来实现本申请实施例的电子设备（例如图 1 所示的服务器或终端设备）的计算机装置 900 的结构示意图。图 9 示出的电子设备仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

如图 9 所示，计算机装置 900 包括中央处理单元（CPU）901 和图形处理器（GPU）902，其可以根据存储在只读存储器（ROM）903 中的程序或者从存储部分 909 加载到随机访问存储器（RAM）904 中的程序而执行各种适当的动作和处理。在 RAM 904 中，还存储有计算机装置 900 操作所需的各种程序和数据。CPU 901、GPU902、ROM 903 以及 RAM 904 通过总线 905 彼此相连。输入/输出（I/O）接口 906 也连接至总线 905。

以下部件连接至 I/O 接口 906：包括键盘、鼠标等的输入部分 907；包括诸如、液晶显示器（LCD）等以及扬声器等的输出部分 908；包括硬盘等的存储部分 909；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 910。通信部分 910 经由诸如因特网的网络执行通信处理。驱动器 911 也可以根据需要连接至 I/O 接口 906。可拆卸介质 912，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 911 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 909。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分 910 从网络上被下载和安装，和/或从可拆卸介质 912 被安装。在该计算机程序被中央处理单元（CPU）901 和图形处理器（GPU）902 执行时，执行本申请的方法中限定的上述功能。

需要说明的是，本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读介质或者是上述两者的任意组合。计算机可读介质例如可以是——但不限于——电、磁、光、电磁、红外线或半导体的装置、装置或器件，或者任意以上的组合。计算机可读介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件或者上述的任意合适的组合。在本申请中，计算机可读介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行装置、装置或者器件使用或者与其结合使用。而在本申请中，计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行装置、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF 等等，或者上述的任意合适的组合。

可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部

分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，也可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本申请各种实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分，该模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的装置来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的模块可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

工业实用性

本发明一种基于混合特征交互 Transformer 的单帧图像超分辨率方法及装置，通过其提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；将低分辨率图像输入经训练的单帧图像超分辨率模型，通过浅层特征提取单元提取浅层特征，将浅层特征输入深层特征提取单元提取得到深层特征，将深层特征输入上采样重建单元，重建得到高分辨率重建图像，解决 Transformer 的 SR 方法忽略不同维度的特征之间的潜在关联性，影响重建性能的问题。

权 利 要 求

1.一种基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，包括以下步骤：

获取待重建的低分辨率图像；

构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型，所述单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元，所述深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；

将所述待重建的低分辨率图像输入所述经训练的单帧图像超分辨率模型，通过所述浅层特征提取单元提取浅层特征，将所述浅层特征输入所述深层特征提取单元提取得到深层特征，将所述深层特征输入所述上采样重建单元，重建得到高分辨率重建图像。

2.根据权利要求 1 所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，所述混合特征交互 Transformer 模块包括高效局部特征提取单元、第一层标准化层、混合特征交互自注意力单元、第二层标准化层和混合尺度前馈神经网络，所述混合特征交互 Transformer 模块的输入依次经过所述高效局部特征提取单元和第一层标准化层，所述第一层标准化层的输出与所述混合特征交互 Transformer 模块的输入相加，得到第一混合特征，所述第一混合特征依次经过所述混合特征交互自注意力单元和第二层标准化层，所述第二层标准化层的输出与所述第一混合特征相加，得到第二混合特征，所述第二混合特征输入所述混合尺度前馈神经网络，得到所述混合特征交互 Transformer 模块的输出。

3.根据权利要求 2 所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，所述高效局部特征提取单元包括依次连接的第一位移卷积层、第一 GeLU 激活函数层、第二位移卷积层、SE 模块和第三位移卷积层，所述高效局部特征提取单元的计算过程如下：

$$H_{ELF}(\cdot) = F_{shift-conv} \left(F_{SE} \left(F_{shift-conv} \left(GeLU \left(F_{shift-conv}(\cdot) \right) \right) \right) \right)$$

；

其中， $H_{ELF}(\cdot)$ 表示高效局部特征提取单元的函数， $F_{shift-conv}(\cdot)$ 表示第

一位移卷积层、第二位移卷积层或第三位移卷积层的位移卷积操作， $F_{SE}(\cdot)$ 表示 SE 模块的函数， $GeLU(\cdot)$ 表示 GeLU 激活函数。

4.根据权利要求 2 所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，所述混合特征交互自注意力单元包括局部窗口自注意力分支、深度卷积分支和双向特征交互单元，所述双向特征交互单元包括空间交互单元和通道交互单元，所述通道交互单元包括依次连接的全局平均池化层、第一卷积层、第一批标准化层，第二 GeLU 激活函数层、第二卷积层和第一 Sigmoid 激活函数层，所述空间交互单元包括依次连接的第三卷积层、第二批标准化层、第三 GeLU 激活函数层、第四卷积层和第二 Sigmoid 激活函数，所述局部窗口自注意力分支包括查询线性层、键线性层、值线性层和局部窗口自注意力模块，所述深度卷积分支包括卷积核大小为 3×3 的第一深度卷积层，所述第一深度卷积层输出的局部特征输入所述通道交互单元，得到通道级动态权重，所述通道级动态权重输入所述局部窗口自注意力分支中对所述值线性层输出的值特征图进行自适应校正；所述局部窗口自注意力模块输出的全局特征输入所述空间交互单元，得到空间级动态权重，所述空间级动态权重输入所述深度卷积分支中对所述局部特征进行自适应校正。

5.根据权利要求 4 所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，所述混合特征交互自注意力单元的计算过程如下：

将输入所述混合特征交互自注意力单元的第一特征图 $X \in \mathbb{R}^{C \times H \times W}$ 输入所述

第一深度卷积层，得到所述局部特征 $F_{local} \in \mathbb{R}^{C \times H \times W}$ ，其中， $\mathbb{R}$ 表示实数多维空间， C 、 H 、 W 分别表示所述第一特征图的通道数量、长度和宽度， $\mathbb{R}^{C \times H \times W}$ 表示形状为 $C \times H \times W$ 的三维数据，窗口尺寸为 S ，其表达式如下：

$$F_{local} = DwConv_{3 \times 3}(X);$$

其中， $DwConv_{3 \times 3}(\cdot)$ 表示第一深度卷积层的函数；

将所述局部特征 F_{local} 输入通道交互单元，获得通道级动态权重

$W_{ca} \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$W_{ca} = CI(F_{local}) ;$$

其中， $CI(\cdot)$ 表示通道交互单元的函数；

将所述第一特征图 X 分割为 N 个窗口尺寸为 $S \times S$ 的非重叠窗口

$X_{win} \in \mathbb{R}^{NS^2 \times C}$ ，其中 $N = H \times W / S^2$ ， $\mathbb{R}^{NS^2 \times C}$ 表示形状为

$NS^2 \times C$ 的二维数据，分别通过所述查询线性层、键线性层、值线性层将

所述非重叠窗口 X_{win} 分别转换为查询特征图 $Q \in \mathbb{R}^{NS^2 \times C}$ 、键特征图

$K \in \mathbb{R}^{NS^2 \times C}$ 和值特征图 $V \in \mathbb{R}^{NS^2 \times C}$ ，其表达式如下：

$$Q, K, V = L_Q(X_{win}), L_K(X_{win}), L_V(X_{win}) ;$$

其中， L_Q 、 L_K 、 L_V 分别表示查询线性层、键线性层、值线性层的函数；

将所述值特征图 V 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ 并与

所述通道级动态权重 W_{ca} 相乘，以对所述值特征图 V 进行自适应校正，随后再

将数据格式还原为 $NS^2 \times C$ ，校正后的结果记作 V' ；

执行局部窗口自注意力模块的计算，以获得全局特征

$F_{global} \in \mathbb{R}^{NS^2 \times C}$ ，表达式如下：

$$F_{global} = \text{Softmax} \left(\frac{QK^T}{\sqrt{C}} V' \right);$$

其中，T 表示转置矩阵，Softmax 表示 Softmax 函数；

将所述全局特征 F_{global} 的数据格式由 $NS^2 \times C$ 转换为 $C \times H \times W$ ，并输入所述空间交互单元，以获得空间级动态权重 $W_{sa} \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$W_{sa} = SI(F_{global});$$

其中， $SI(\cdot)$ 表示空间交互单元的函数；

通过将所述空间级动态权重 W_{sa} 与全局特征 F_{local} 相乘，以对全局特征 F_{local} 进行自适应校正，其校正后的结果记作 F'_{local} ；

最后，将所述全局特征 F_{local} 与 F'_{local} 相加，以获得混合特征

$$F_{mix} \in \mathbb{R}^{C \times H \times W}。$$

6.根据权利要求2所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法，其特征在于，所述混合尺度前馈神经网络包括第一分支、第二分支和第五卷积层，所述第一分支包括依次连接的第二深度卷积层、第一 ReLU 激活函数层、第三深度卷积层和第二 ReLU 激活函数层，所述第二分支包括依次连接的第四深度卷积层、第三 ReLU 激活函数层、第五深度卷积层和第四 ReLU 激活函数层，其中所述第二深度卷积层和第五深度卷积层的卷积核大小为 7×7 ，所述第三深度卷积层和第四深度卷积层的卷积核大小为 5×5 ，具体计算过程如下：

沿着通道维度将输入所述混合尺度前馈神经网络的第二特征图

$X' \in \mathbb{R}^{C \times H \times W}$ 将 X' 切分为均等的两份，得到切分后的特征

$X_{slice}^1 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 和 $X_{slice}^2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ ， $\mathbb{R}^{\frac{C}{2} \times H \times W}$ 表示形状为

$\frac{C}{2} \times H \times W$

的三维数据，将 X_{slice}^1 和 X_{slice}^2 分别输入所述第一分支和第二分支进行混合交叉特征提取，分别输出第一交叉特征和第二交叉特征，将所述第一交叉特征和第二交叉特征进行拼接后输入所述第五卷积层，所述第五卷积

层的输出与所述第二特征图相加，得到混合尺度特征 $Y \in \mathbb{R}^{C \times H \times W}$ ，其表达式如下：

$$X_1^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_{slice}^1\right)\right);$$

$$X_1^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_{slice}^2\right)\right);$$

$$X_2^{p_1} = \text{ReLU}\left(\text{DwConv}_{5 \times 5}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$X_2^{p_2} = \text{ReLU}\left(\text{DwConv}_{7 \times 7}\left(X_1^{p_1} + X_1^{p_2}\right)\right);$$

$$Y = \text{Conv}_{1 \times 1}\left([X_2^{p_1}, X_2^{p_2}]\right) + X';$$

其中， $\text{ReLU}(\cdot)$ 表示 ReLU 激活函数， $\text{DwConv}_{5 \times 5}(\cdot)$ 和

$\text{DwConv}_{7 \times 7}(\cdot)$ 分别表示卷积核为 5×5 和 7×7 的深度卷积层的函数，

$\text{Conv}_{1 \times 1}(\cdot)$ 表示卷积核大小为 1×1 的第五卷积层的函数， $[\cdot]$ 表示拼接操

作, $X_1^{p_1}$ 、 $X_1^{p_2}$ 分别表示第一特征和第二特征, $X_2^{p_1}$ 、 $X_2^{p_2}$ 分别表示第一交叉特征和第二交叉特征。

7.根据权利要求 1 所述的基于混合特征交互 Transformer 的单帧图像超分辨率方法, 其特征在于, 所述单帧图像超分辨率模型的具体结构和计算过程如下:

所述浅层特征提取单元采用第六卷积层, 所述浅层特征提取单元的计算过程如下:

$$F_0 = \text{Conv}_{3 \times 3}(I_{LR});$$

其中, F_0 表示浅层特征 $F_0 \in \mathbb{R}^{C \times H \times W}$, $\text{Conv}_{3 \times 3}(\cdot)$ 表示卷积核为

3×3 的第六卷积层的函数, I_{LR} 表示低分辨率图像;

利用 P 个所述混合特征交互 Transformer 模块提取特征, 并利用长跳跃连接将 F_0 传递至网络末端, 与第 P 个混合特征交互 Transformer 模块的输出相加以进行残差学习, 其表达式如下:

$$F_i = \text{MF}^i(F_{i-1}), i \in [1, P];$$

$$F_{P0} = \text{MF}^P(\dots(\text{MF}^2(\text{MF}^1(F_0)))) + F_0;$$

其中, F_{i-1} 表示第 $i-1$ 个所述混合特征交互 Transformer 模块的输出, MF^P 表示第 P 个所述混合特征交互 Transformer 模块的函数, MF^1 表示第 1 个所述混合特征交互 Transformer 模块的函数, MF^2 表示第 2 个所述混合特征交互 Transformer

模块的函数, MF^i 表示第 i 个所述混合特征交互 Transformer 模块的函数, F_i 表

示第 i 个所述混合特征交互 Transformer 模块的输出, F_{P0} 表示深层特征,

$$F_{P0} \in \mathbb{R}^{C \times H \times W};$$

所述上采样重建单元包括尺度因子为 scale 的亚像素卷积层和卷积核为 3×3

的第七卷积层，其表达式如下：

$$I_{SR} = \text{Conv}_{3 \times 3} \left(f_{ip} (F_{P0}) \right);$$

其中， $f_{ip}(\cdot)$ 表示亚像素卷积层的函数， $\text{Conv}_{3 \times 3}(\cdot)$ 表示第七卷积层的函数， I_{SR} 表示高分辨率重建图像， $I_{SR} \in \mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$ ，

$\mathbb{R}^{3 \times (H \times scale) \times (W \times scale)}$

表示形状为 $3 \times (H \times scale) \times (W \times scale)$ 的三维数据。

8.一种基于混合特征交互 Transformer 的单帧图像超分辨率装置，应用权利要求 1-7 中任一所述的方法,其特征在于，包括：

图像获取模块，被配置为获取待重建的低分辨率图像；

模型构建模块，被配置为构建基于混合特征交互 Transformer 的单帧图像超分辨率模型并训练，得到经训练的单帧图像超分辨率模型，所述单帧图像超分辨率模型包括依次连接的浅层特征提取单元、深层特征提取单元和上采样重建单元，所述深层特征提取单元包括依次连接的 P 个混合特征交互 Transformer 模块；

重建模块，被配置为将所述待重建的低分辨率图像输入所述经训练的单帧图像超分辨率模型，通过所述浅层特征提取单元提取浅层特征，将所述浅层特征输入所述深层特征提取单元提取得到深层特征，将所述深层特征输入所述上采样重建单元，重建得到高分辨率重建图像。

9.根据权利要求 8 所述的基于混合特征交互 Transformer 的单帧图像超分辨率装置，其特征在于，所述混合特征交互 Transformer 模块包括高效局部特征提取单元、第一层标准化层、混合特征交互自注意力单元、第二层标准化层和混合尺度前馈神经网络。

10.一种电子设备，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-7 中任一所述的方法。

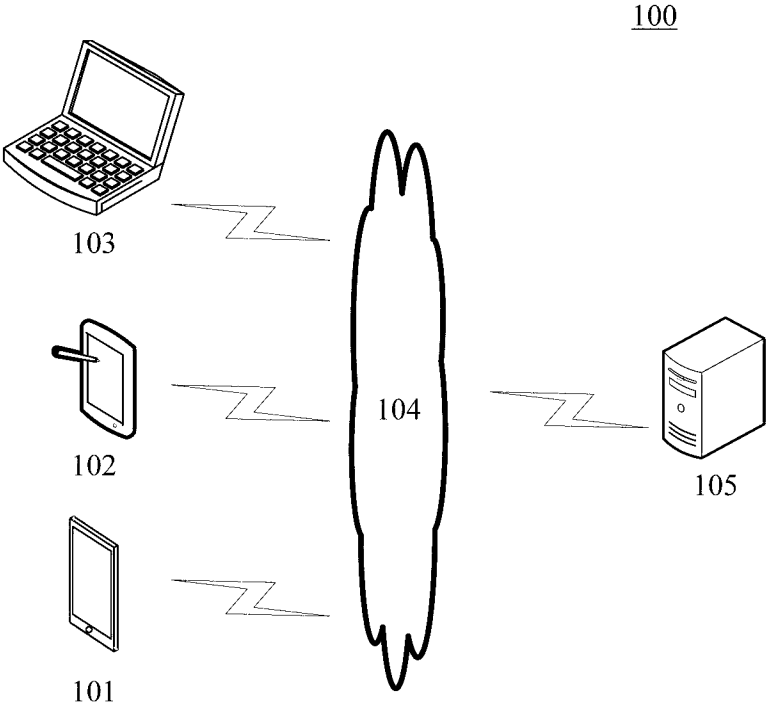


图 1

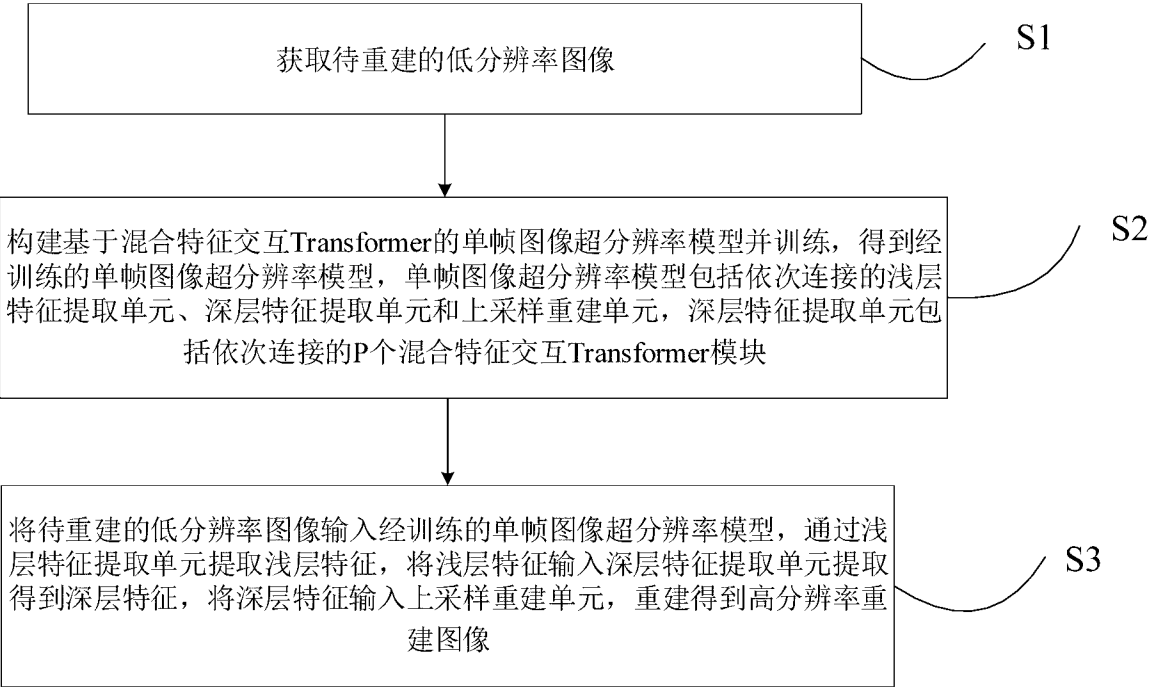


图 2

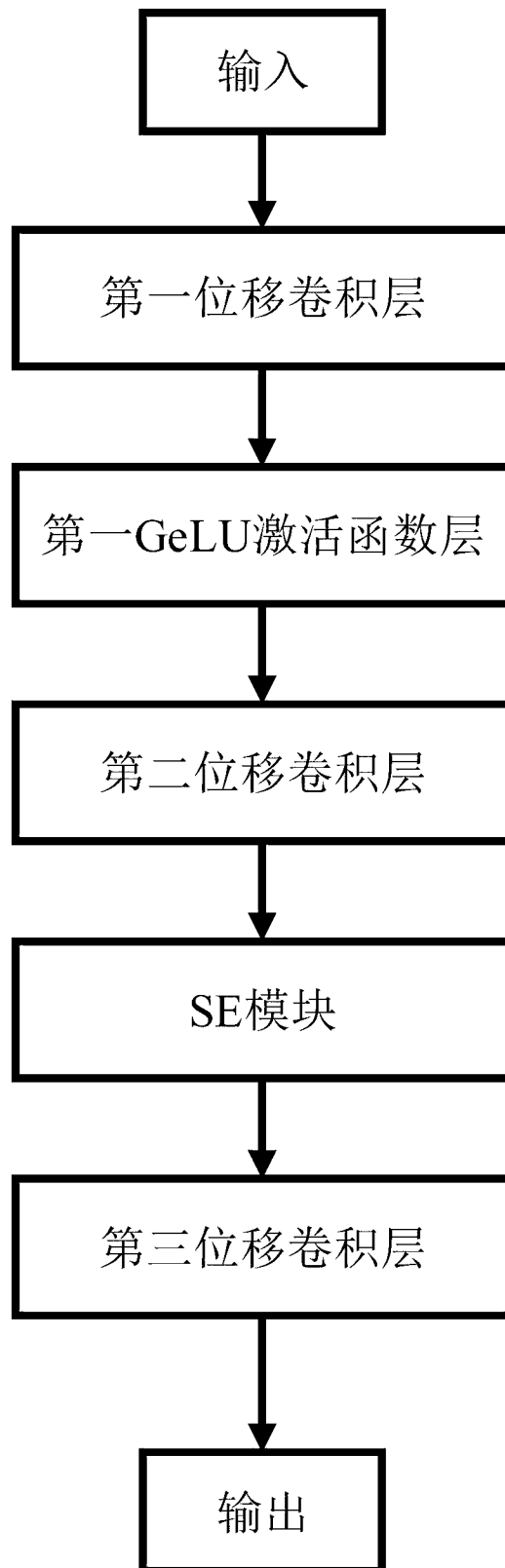


图 3

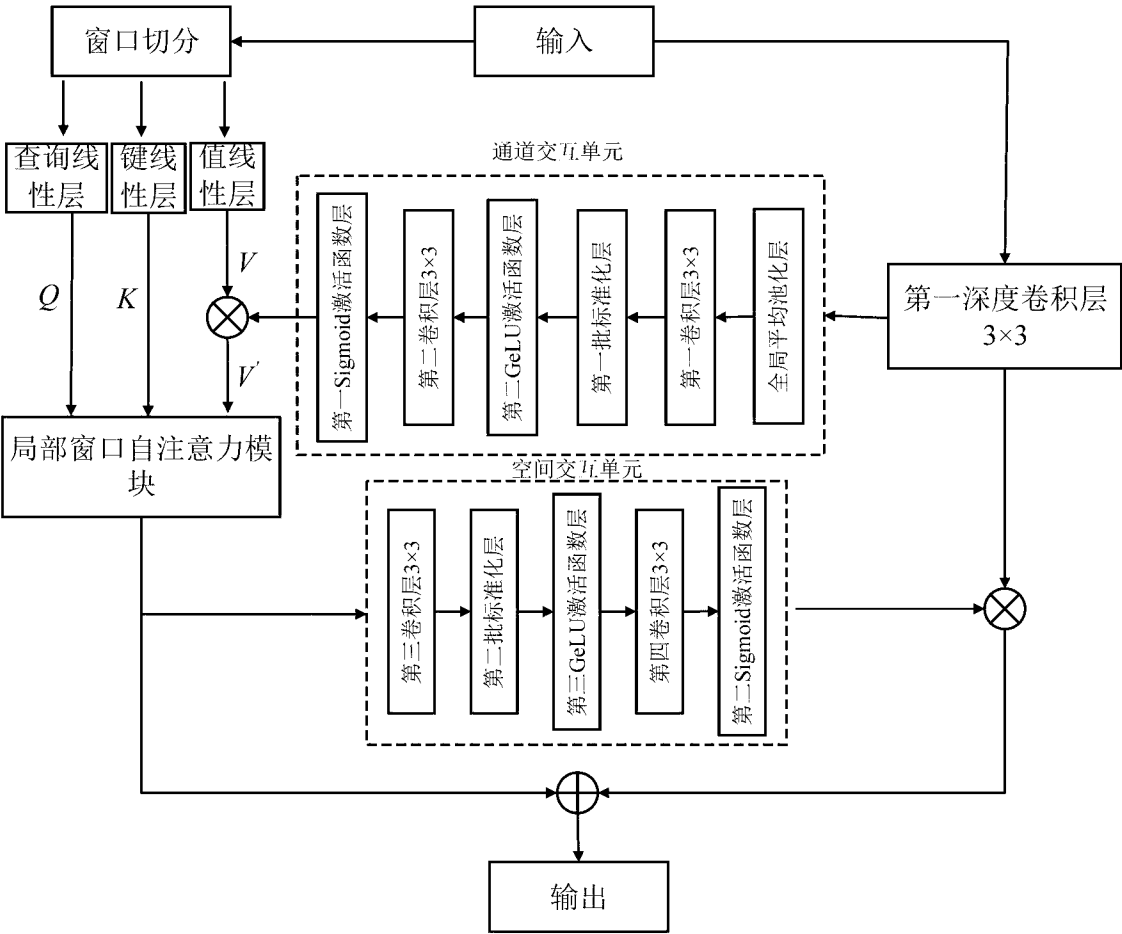


图 4

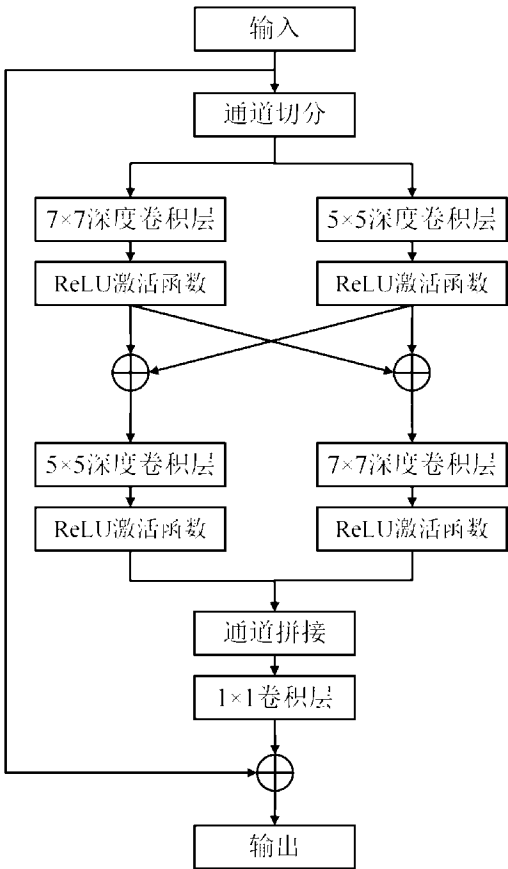


图 5

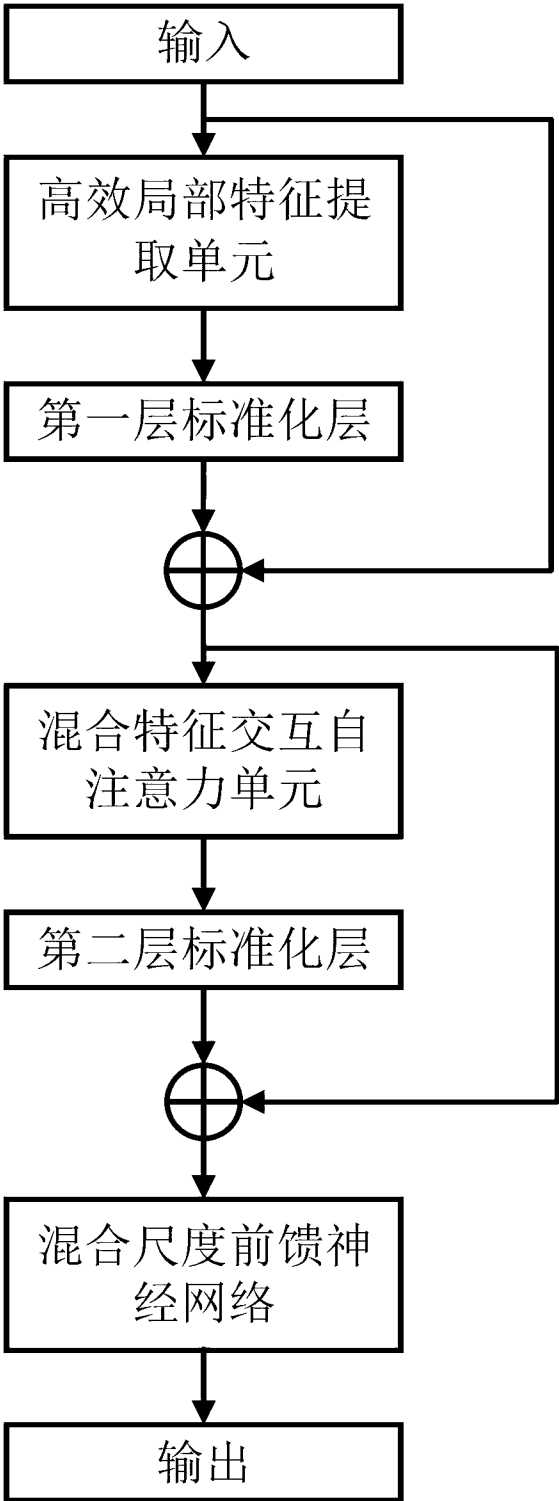


图 6

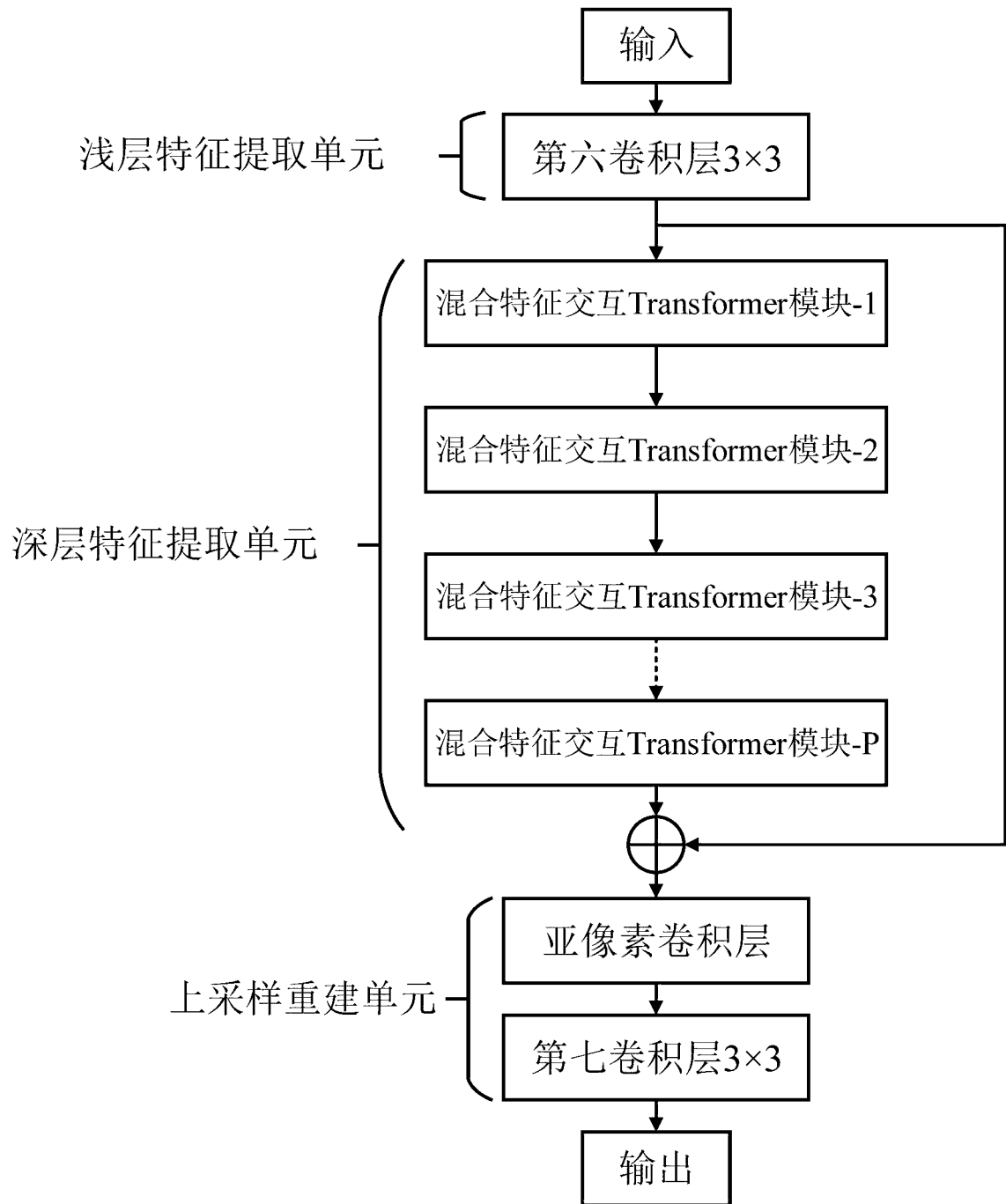


图 7

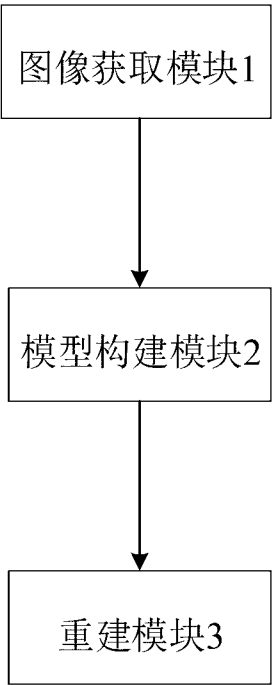


图 8

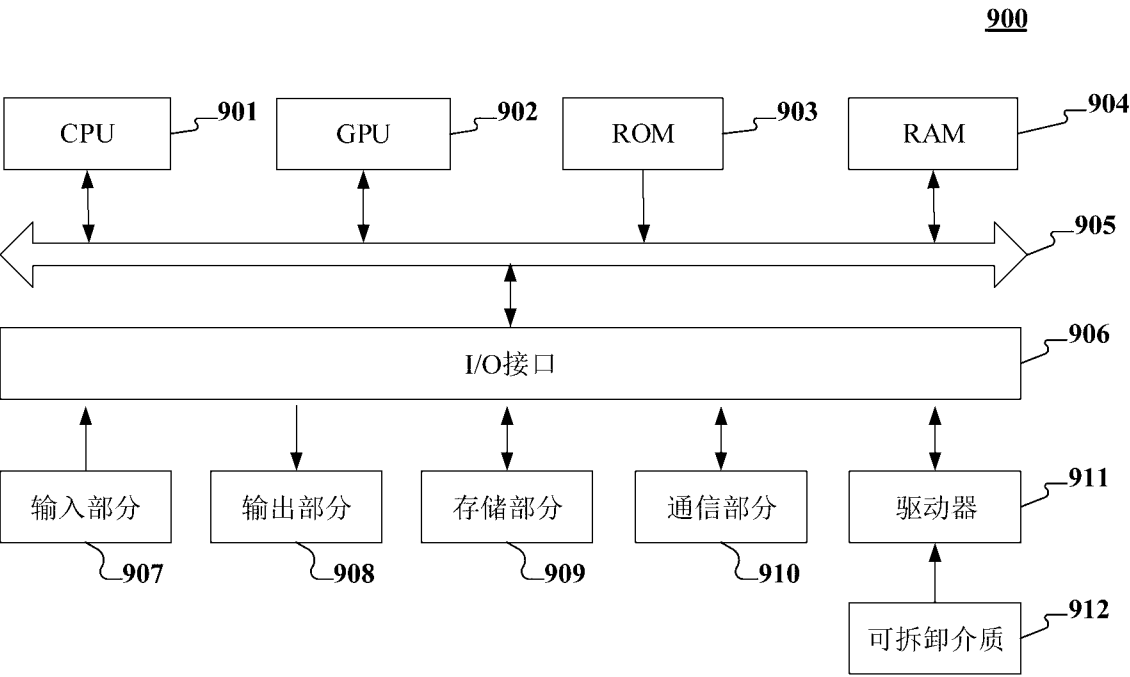


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/143017

A. CLASSIFICATION OF SUBJECT MATTER

G06T3/40(2024.01)i; G06N3/0455(2023.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06T G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, WPABS, DWPI, USTXT, WOTXT, EPTXT, CNKI, IEEE, Web of Science, 百度, BAIDU, 万方, WANFANG: 混合, 特征, 交互, transformer, 单帧, 单幅, 图像, 超分辨率, 重建, 浅层, 深层, 自注意力, 卷积, hybrid, feature, interaction, single, frame, image, super resolution, reconstruction, shallow, deep layer, self-attention, convolution

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 114926337 A (MINJIANG UNIVERSITY) 19 August 2022 (2022-08-19) description, paragraphs [0005]-[0038]	1, 8-10
A	CN 114387161 A (SICHUAN UNIVERSITY) 22 April 2022 (2022-04-22) entire document	1-10
A	CN 117237197 A (HUAQIAO UNIVERSITY) 15 December 2023 (2023-12-15) entire document	1-10
A	CN 117173025 A (HUAQIAO UNIVERSITY) 05 December 2023 (2023-12-05) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

03 August 2024

Date of mailing of the international search report

26 August 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/143017

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114926337	A	19 August 2022	None			
CN	114387161	A	22 April 2022	CN	114387161	B	07 July 2023
CN	117237197	A	15 December 2023	CN	117237197	B	01 March 2024
CN	117173025	A	05 December 2023	CN	117173025	B	01 March 2024

A. 主题的分类

G06T3/40(2024.01)i; G06N3/0455(2023.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: G06T G06N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;WPABS;DWPI;USTXT;WOTXT;EPTXT;CNKI;IEEE;Web of Science;百度;万方: 混合, 特征, 交互, transformer, 单帧, 单幅, 图像, 超分辨率, 重建, 浅层, 深层, 自注意力, 卷积, hybrid, feature, interaction, single, frame, image, super resolution, reconstruction, shallow, deep layer, self-attention, convolution

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 114926337 A (闽江学院) 2022年8月19日 (2022 - 08 - 19) 说明书第[0005]-[0038]段	1、8-10
A	CN 114387161 A (四川大学) 2022年4月22日 (2022 - 04 - 22) 全文	1-10
A	CN 117237197 A (华侨大学) 2023年12月15日 (2023 - 12 - 15) 全文	1-10
A	CN 117173025 A (华侨大学) 2023年12月5日 (2023 - 12 - 05) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

★ 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“p” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2024年8月3日

国际检索报告邮寄日期

2024年8月26日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

于俊

电话号码 (+86) 0512-88995917

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/143017

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	114926337	A	2022年8月19日	无	
CN	114387161	A	2022年4月22日	CN	114387161 B 2023年7月7日
CN	117237197	A	2023年12月15日	CN	117237197 B 2024年3月1日
CN	117173025	A	2023年12月5日	CN	117173025 B 2024年3月1日