Rethinking Perceptual Masking Integration in JND Modeling: An Efficient and Explainable Learning Approach

摘要

1、引言

Traditional JND models often rely on simple linear and nonlinear fusion techniques that fail to accurately represent the complex mechanisms of the Human Visual System (HVS). To address these limitations, we propose a new perceptual masking integration model that leverages neural architecture search (NAS) to automatically discover optimal fusion strategies. Our approach introduces interpretable feature mapping units and a flexible framework for organizing and selecting masking operations, making the model both efficient and explainable. Extensive experiments demonstrate that our method significantly improves the accuracy of JND predictions compared to existing methods. The proposed model is particularly effective in applications such as image compression, where it provides better visual quality and bitrate savings. Furthermore, cross-database validation confirms the generalizability of our approach, making it a robust solution for various image processing tasks

.2、相关工作

由于视觉 JND 是人类视觉感知现象的结果，这一认知启示我多媒体社区基于相关生理学、心理学和脑科学的最新研究成果制定计算模型，以提取视觉感知相关的视觉掩蔽特征。因此，我们根据视觉掩蔽特征的计算方法将现有方法分为基于手工的方法和基于学习的方法两大类。此外，由于人类视觉感知是一个多因素问题，现有 JND 模型在计算过程中不可避免地需要将提取到的视觉掩蔽特征进行融合。因此，我们根据融合方法的不同，对现有方法进一步分类。具体地，基于手工视觉掩蔽效应的融合主要包含线性和非线性融合，基于深度视觉掩蔽效应的融合可以分为压缩融合，加性融合和拼接融合。我们在接下来的内容中对每类方法做了更详细的描述。

2.1、基于手工视觉掩蔽效应的融合

目前，基于手工视觉效应的方法通过研究人类视觉系统(Human Visual System, HVS)的生理和心理视觉效应进行开发。这些方法对 HVS 中的各种视觉效应（诸如 LA，CM）进行手工建模，并采用特定的融合方法整合这些视觉效应以预测视觉冗余。根据融合

方法的不同，现有基于手工视觉效应的融合方法可以分为线性融合和非线性融合两类。

线性融合（Linear Integration）主要采用线性运算（诸如加权，乘法）融合视觉效应。线性融合中一种常见方法是加权融合，即根据视觉效应的重要程度经验性为视觉效应设置权重。例如亮度掩蔽和时空掩蔽被经验性地赋予权重，以更好的权衡视觉冗余度量和重建感知质量[1]。类似地，纹理掩蔽被视为边缘掩蔽和纹理掩蔽的加权和以更符合人类视觉系统的熵掩蔽特性[2]。乘法融合是另一种常用方法，该方法将视觉效应作为调整因子与基础阈值直接相乘以修正视觉阈值。为了建立更准确的视觉冗余模型，近期的方法基于乘法融合方法提出了LM[3]和CM[4]的修正模型，并引入了更多的调整因子，如时序特征[5]，颜色特征[6]，变换块大小[7]，方向敏感性[8]等。

非线性融合（Non-linear Integration）引入了最大值，最小值这一类非线性计算单元以整合视觉效应。早期的方法直接采用非线性单元融合视觉效应，例如，视觉冗余估计被简化为LA和CM的主导效应[9]，掩蔽效应被视为对比掩蔽和模式掩蔽的最大强度[10]，亮度和色度通道的最小视觉阈值被作为整体的视觉阈值[11]等。近期的方法将视觉效应的复合作用被视为叠加作用，并提出非线性加性模型（Non-linear Additivity Model for Masking，NAMM）以去除视觉效应之间的重叠效应[12]。基于NAMM这一融合方法，现有方法引入了更多视觉效应（诸如恰可察觉模糊[8]，模式复杂度[10]，无序掩蔽[13]）和更多感知先验知识（诸如多域感知结合[14]，内容区域划分[15]，层级预测编码[16]）。

然而，尽管人类视觉系统的感知特性已经得到了广泛研究，但如何整合视觉效应还处于初步研究阶段。基于手工视觉效应的融合方法至少存在以下三个问题：1）现有方法均采用线性和非线性融合方法，这类简单的融合方法难以准确拟合HVS这一高度复杂系统，这限制了整体模型的预测准确性；2）上述融合方法运行在单个像素或系数级别，这类方法在融合过程中无法考虑上下文信息，这限制了视觉效应的表征潜力；3）手工制定的融合方法往往需要在特定数据集上进行大量主观实验进行参数设置，这种方法容易受主观偏差和数据特性的影响，从而限制融合方法的泛化能力。

2.2、基于深度视觉掩蔽效应的融合

基于学习的视觉冗余预测方法能借助于神经网络的高度非线性表征能力能有效地模拟视觉感知，成为近期视觉冗余估计领域内的研究热点。基于学习的方法自动从图像中提取视觉感知特征，并融合提取到的特征以实现视觉冗余预测。现有基于学习的方法根据其融合方法不同可以分为：压缩融合，加性融合和拼接融合。

压缩融合（Squeeze Integration）的方法旨在将不同通道内的视觉效应特征进行融合，促进通道之间的信息交互以获得更丰富的特征表达。其中，一种常见的方法是使用 1X1 卷积进行通道间的信息交互。1x1 卷积通过压缩特征通道维度，汇聚通道间多种深度视觉特征，以实现通道间的信息融合[8]。另一种常见的方法是采用通道注意力机制动态整合全局通道信息。通道注意力机制通过挤压和激励操作，自适应地整合全局通道信息以学习深度视觉效应特征的相对重要性，并对不同特征进行加权调整[17]。然而，这类方法主要依赖于网络学习能力。网络特征通道数量往往非常大，需要学习大量的参数进行特征融合，网络可能面临训练困难的挑战。

加性融合（Additive Integration）的方法采用线性加法运算直接融合不同特征，调整深度视觉效应特征响应强度以增加特征表达能力。现有的方法往往结合领域知识和先验信息根据对特征的分析，为特征融合选择合适的线性运算。例如，考虑到视觉冗余是对原图和失真图差异的衡量，原图与失真图之间的特征差被用来表征失真程度[18, 19]；基于视觉特征在视觉冗余估计中的相似性和差异性，特征的求和增强和特征的差异偏移被用来挖掘特征间的互补关系[17]。然而，这类直接或启发式的线性运算缺乏对特征之间复杂关系的建模能力，可能无法挖掘特征间的非线性互补性，这限制了融合特征的表征潜力。

拼接融合（Concatenation Integration）的方法是将不同特征直接串联来实现特征融合，旨在同时利用多个特征信息以增加特征的多样性。为了同时利用多样的特征信息，这类方法通常直接将表征不同的特征拼接起来以构建更丰富的特征表示。常用的方法包括在输入阶段拼接额外的视觉效应特征以利用更多先验知识，如亮度信息[19]，视觉显著性[20]，模式复杂度[20]，时序特征[21] 等，或者将早期低级空间特征与晚期高级语

义特征进行跳接以综合利用特征信息[22]等。然而，直接拼接特征增加了特征维度会导致计算量暴增，从而增大网络训练的负担。此外，特征拼接可能引入冗余和噪声信息，从而影响网络性能。

为了解决上述挑战，我们开发了一种通用型感知掩蔽融合方法为视觉特征自动搜索更优的融合网络。此外，我们基于 HVS 视觉感知设计了特征映射微元，并将它们嵌入到网络搜索中，使融合网络对掩蔽效应的表征更符合视觉感知。

3、感知掩蔽融合模型

本文旨在提出一种高效且可解释的感知掩蔽融合模型。在本节中，我们首先基于现有工作的回顾对视觉掩蔽效应融合进行了重新思考，其次我们明确了融合模型的优化目标，最终我们详细介绍了所提出的融合网络搜索方法。图 x 展示了使用本文方法搜索融合网络的例子。

3.1、重新思考掩蔽融合：

视觉掩蔽融合技术旨在将不同视觉掩蔽特征有机地结合，以预测视觉冗余效应。在视觉掩蔽融合过程中，我们主要关注特征映射，融合层级和融合算子三个关键方面。

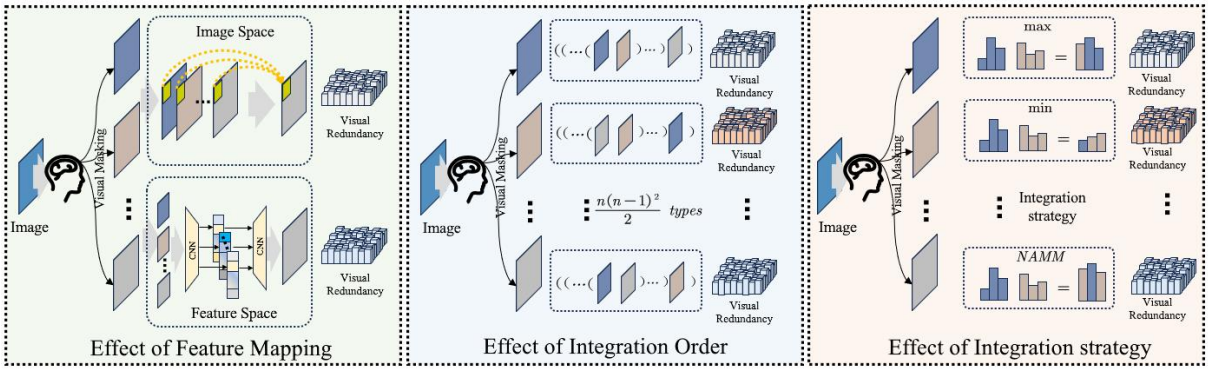


Figure 1 特征映射、融合层级和融合算子对视觉掩蔽效应融合的影响。

3.1.1、特征映射的影响：

特征映射主要目的是将视觉掩蔽特征映射到相似的空间。不同视觉掩蔽特征的建模方法通常具有较大差异，缺乏直接的关联性。为了使得融合过程更顺利，这些异质的视觉掩蔽特征需要被映射到相似的空间。因此，特征映射是视觉掩蔽融合的基础，会影响视觉冗余预测模型的性能。

现有方法根据视觉掩蔽特征的类型设计特征映射方法。基于手工建模的方法从亮度

差异，模糊敏感度等多种角度建模视觉掩蔽特征，并基于主观实验将各种特征统一映射到JND水平直接融合。与手工建模方法不同，基于深度神经网络（DNN）方法将不同视觉掩蔽特征统一映射到特征空间，并通过数据驱动的方式让不同视觉掩蔽特征在特征空间的分布尽可能相似，以促进特征融合。然而，上述特征映射方法未考虑HVS的感知特性，且缺乏可解释性。

因此，在特征融合时需要考虑特征映射空间的影响。合适的特征映射能将视觉掩蔽特征映射到更相似的映射空间，使得多种特征的融合过程更自然，从而提高视觉冗余预测的准确性。

为了研究视觉掩蔽特征映射空间对视觉冗余估计的影响，我们基于LA，CM两种视觉掩蔽特征，分别采用图像空间映射和特征空间映射在基准数据库CSIQ上进行视觉冗余预测对比实验。具体的，图像空间映射方式我们采用NAMM方式融合LA和CM，特征空间映射方式我们采用两层普通卷积融合LA和CM。实验中固定注入噪声水平为MSE=100，评估指标为MS-SSIM，实验结果如下图x所示。从实验结果可以看出，模型性能与特征映射方式相关。因此，选择合适的掩蔽特征映射方式对提高视觉冗余预测准确性很重要。

3.3.2、掩蔽次序的影响：

掩蔽次序是指视觉掩蔽特征在融合过程中的先后顺序。HVS是一个层级系统，不同的视觉层级处理着不同的视觉信号，从而导致视觉掩蔽特征具有不同的融合次序。\cite{Wang2020Hierarchical}。这种分层结构赋予了HVS出色的感知能力，使我们能够感知到丰富多样的视觉信息。因此，掩蔽次序在特征融合中扮演着核心角色，会对最终的视觉冗余预测准确度产生影响。

现有的方法往往会基于经验来设计掩蔽次序。常见的做法包括先融合亮度自适应等基础视觉掩蔽特征，或者根据视觉内容特点分区域融合视觉掩蔽特征，或者一次融合所有视觉掩蔽特征等。然而，现有融合方法均是根据特定的视觉掩蔽特征或应用场景设计，难以在更普遍的场景下使用，缺乏可扩展性。

因此，在融合过程中对视觉掩蔽特征的掩蔽次序进行灵活的扩展和调整以适应不同图像类型和应用场景的需求，这将有助于进一步提升视觉冗余估计的性能，并为视觉冗余的应用提供更可靠的支持。

基于上述讨论，我们进了一个对比实验以研究融合层级对视觉冗余估计的影响。

具体地，我们对Wu2013TMM[]方法中三种视觉效应的三种融合层级情况在基准数据库CSIQ上做了对比实验。实验中固定注入噪声水平为MSE=100，评估指标为MS-SSIM，实验结果如下图所示。从实验结果可以看出，融合层级在不同融合层级下具有不同的性能。因此，为了提高整体模型的性能，寻找合适的融合层级很有必要。

3.1.3、融合算子的影响：

融合算子的主要目标是模拟HVS对多种视觉掩蔽特征的联合感知机制，在视觉掩蔽特征融合中扮演着重要角色。HVS的视觉冗余是多种视觉掩蔽特征相互作用的结果，这一感知机制是一个复杂而精密的过程，它使我们能够高效地感知并理解复杂的视觉信息\cite{Lin2022Progress}。因此，融合算子在掩蔽特征融合的关键，会影响视觉冗余预测性能。

现有方法根据视觉掩蔽特征的相互关联设计融合算子，主要可以分为显式和隐式两种。其中，显式融合算子通常基于领域内的专业知识分析特征之间的关系，并根据HVS的感知特性为视觉掩蔽特征设计确定的融合算子。另一方面，由于HVS中不同因素融合的确切机制尚未得到充分研究，一些基于学习的方法采用隐式融合，也即不给定明确的融合算子，期望神经网络自适应学习HVS对视觉掩蔽特征的综合感知机制。然而，现有的融合算子均是针对特定的特征设计，难以应用于其他类型视觉掩蔽特征融合，缺乏通用性。

因此，合适的融合算子能使得掩蔽特征融合过程更符合HVS感知机制。此外，优秀的融合算子需要考虑视觉掩蔽特征之间的联系，以高效且充分地利用特征信息，从而提高视觉冗余预测的准确性。

为了研究融合算子对视觉冗余估计的影响，我们采用不同融合算子进行了对比实验。具体地，我们基于LA和CM分别在最大，最小和NAMM的融合算子下进行对比实验。测试在基准数据库CSIQ上进行，实验中固定注入噪声水平MSE=100，评估指标为MS-SSIM，实验结果如下表所示。从实验结果中可以发现，融合算子能对视觉冗余预测模型产生显著影响。因此，研究如何选择合适的融合算子对视觉冗余预测具有重要意义。

3.2、融合模型优化目标：

3.2.1、问题定义

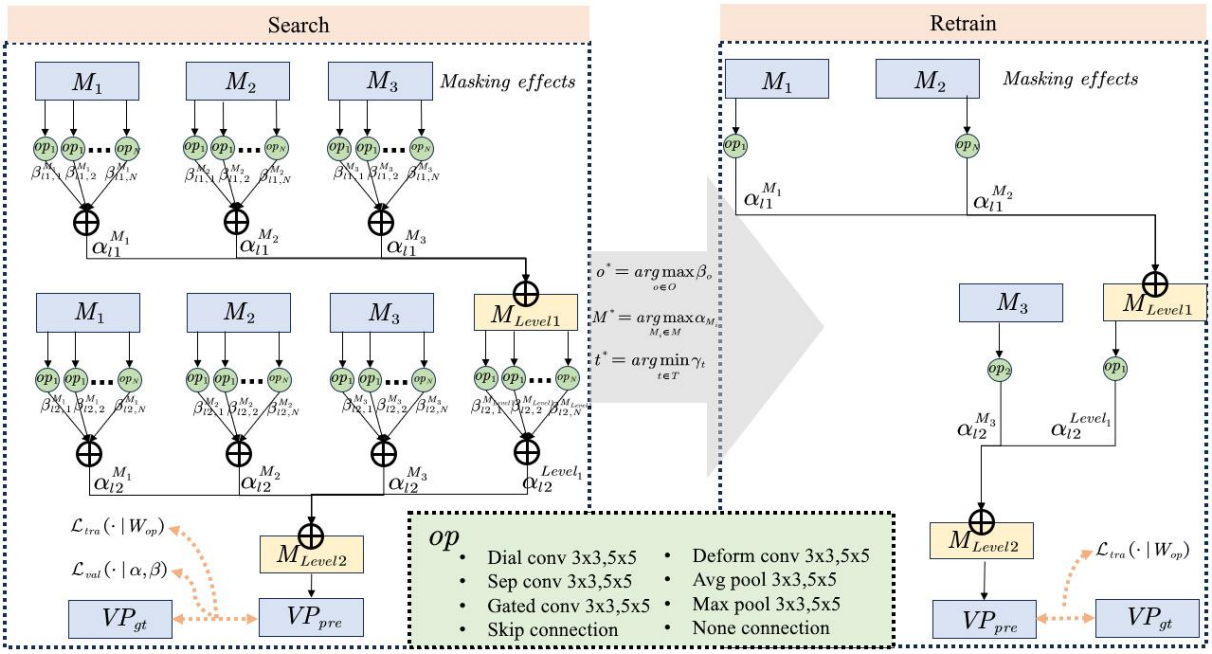


Figure 2 使用提出的方法搜索视觉掩蔽效应融合网络的整理框架。左边是超网络的优化训练过程，右边是根据超网络推理得到的融合网络。

如前所述，我们的目标是为给定视觉特征的寻找合适的融合网络，以尽可能准确地预测视觉冗余。搜索融合网络最直接的方式是测试所有可能的融合结构，根据每种网络性能挑选出最优的融合网络。然而，这种方法的计算成本是难以忍受的。因此，为了更高效地搜索融合网络，我们采用神经架构搜索的方法，可以表示为：

$$S^* = \arg\min_{S} \mathcal{L}(S_{hyper}(V_m), V_{gt}),$$

其中V_m, V_{gt}分别是输入的掩蔽效应和视觉冗余标签，S_{hyper}, S^*分别表示包含所有可能候选网络的超网络和搜索到的最优融合网络。$\mathcal{L}(\cdot)$是预测值和标签的损失函数。

3.2.2、优化目标

为了更高效地进行网络搜索，我们采用了DARTS的持续松弛方案，通过在搜索空间中引入可学习的超参数将搜索空间松弛为连续的，并使用梯度优化方法来搜索最优的网络架构。图x展示了融合3种掩蔽特征的整体优化过程。所提出的感知掩蔽融合网络搜索主要可以分为映射微元选择，掩蔽次序整理和融合算子选择三部分，我们将感

知掩蔽融合网络搜索的主要机制描述如下。

映射微元选择目标是将每种给定掩蔽特征映射到相似的特征空间。为此，我们基于HVS的多种感知特点设计了可解释的特征映射微元候选集$O=\{o_i\}_{i\in[1,N]}$以供搜索，期望能将特征映射到和HVS感知更相似的空间进行融合。每条微元的映射结果均采用超参数β加权。超参数β和所有映射网络层的网络参数W在训练中采用梯度优化的方式进行更新以实现映射微元选择。

掩蔽次序整理旨在搜索给定视觉掩蔽特征的最优融合结构。假设输入n种视觉掩蔽特征$V=\{M_1,M_2,M_3,...,M_n\}$。融合骨架是输入的掩蔽特征和特征n-1个融合层进行有序连接组成的有向无环图，每个有向边都添加了可学习的超参数α，在训练中通过梯度优化的方式进行更新以实现掩蔽次序整理。

融合算子选择旨在为视觉掩蔽特征之间选择最优的融合运算。为此，我们综合了现有方法，预设多种融合算子用于不同视觉掩蔽特征融合的搜索。每种融合算子的结果都被设置了可学习的超参数γ。超参数γ在训练中通过梯度优化的方式进行更新以实现融合算子选择。

我们期望视觉特征经过我们搜索的融合网络尽可能接近视觉阈值。感知掩蔽融合网络搜索的优化问题可表示为如下的优化问题：

$$\begin{aligned}&\underset{\theta_\alpha,\theta_\beta,\theta_\gamma}{arg\min}\mathcal{L}_{val}(S_{hyper}(V_m\mid\theta_W^*),V_{gt}\mid\theta_W^*,\theta_\alpha,\theta_\beta,\theta_\gamma)\\&s.t.\ \theta_W^*=\underset{\theta_W}{arg\min}\mathcal{L}_{tra}(S_{hyper}(V_m\mid\theta_W),V_{gt}\mid\theta_W,\theta_\alpha,\theta_\beta,\theta_\gamma)\end{aligned},$$

其中$\mathcal{L}_{tra},\mathcal{L}_{val}$表示在训练集和验证集上的损失函数。在搜索过程中，数据集按照5:5分为训练集和验证集。训练过程更新表征单元的网络权重，验证过程更新网络搜索的超参数权重，两种更新方式交替进行。

3.3、掩蔽融合模型搜索

3.3.1、可解释的特征映射微元

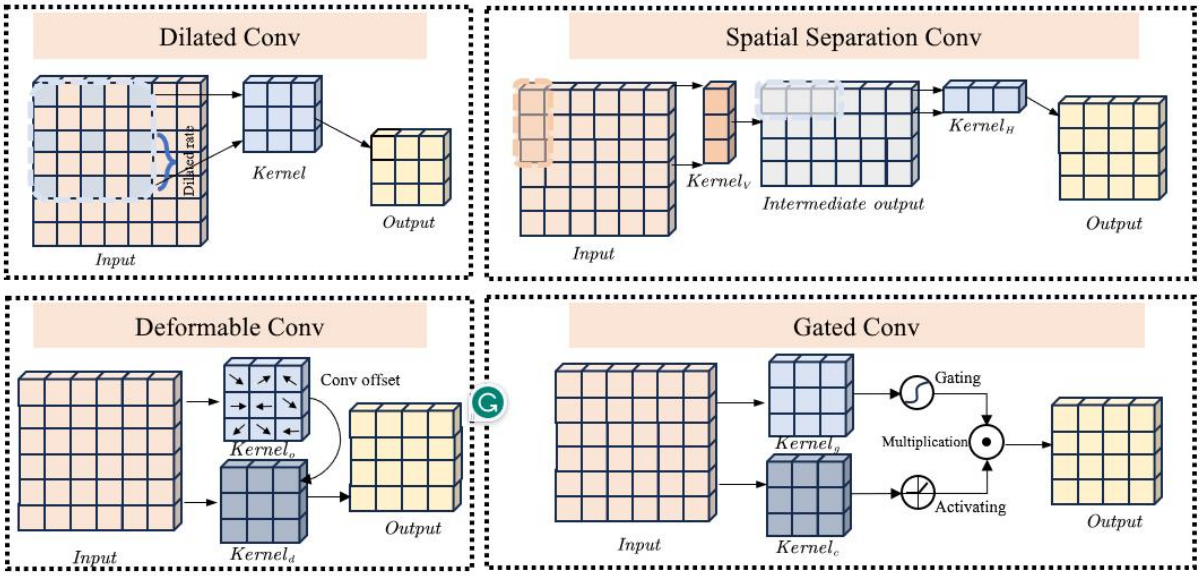


Figure 3 提出的可解释特征映射微元。

如前所述，视觉掩蔽特征的映射空间会影响了融合网络的性能。为了更符合HVS特性地表征各种视觉掩蔽效应，我们根据HVS的感知特点设计了多种特征映射微元组成了搜索空间，具体包括扩张卷积、可形变卷积、门控卷积，空间可分离卷积、最大池化和平均池化。其中每个映射微元采用ReLU-Conv-BatchNormalization的结构：

$$F_o = BN(Conv(\mathrm{Re}LU(F_i))).$$

其中F_i,F_o分别为输入特征和输出特征，ReLU表示矫正线性单元（Rectified Linear Unit），Conv表示任意的特征映射微元，BN为批归一化层（Batch Normalization）。在接下来的内容中，我们将详细描述搜索空间中每种映射单元的设计动机和功能。

扩张卷积通过在卷积核之间引入固定的间隔，能扩大卷积操作中的感受野。人眼视觉系统往往不是孤立的理解像素或区域信息，而是结合上下文环境进行综合感知。扩张卷积中的大感受野能捕捉更广泛和丰富的上下文信息，以便更好地处理图像中上下文信息对视觉感知的影响。扩张卷积可以表示为：

$$F_o(x,y) = \sum_{i,j\in[1,k]} F_i(x+i\times r, y+j\times r)\cdot K(i,j),$$

其中K为卷积核，x,y表示特征所在位置，r为扩张率，扩张率越大，感受野越大。

可形变卷积在卷积核上增加空间偏移量，能自适应调整卷积核的形状和位置来提高图像中几何变换的建模能力。为了更好的理解视觉内容，人眼视觉系统倾向于提取图像中的形状，轮廓，元素交互等模式信息，不同的模式复杂度会导致不同的视觉掩蔽程度。可形变卷积中高效的集合变换建模能力能够充分捕捉图像中的模式信息，以更好地模拟人眼视觉系统对视觉内容的理解过程。可形变卷积可以表示为:

$$F_o(x,y) = \sum_{i,j \in [1,k]} F_i(x + p_h(i,j), y + p_v(i,j)) \cdot K(i,j),$$

其中p_h,p_v分别为特征在水平和垂直方向上的空间偏移量。

门控卷积引入门控机制对卷积进行调控，能根据特征重要性对卷积结果实现动态加权。人眼视觉系统对不同的像素或区域敏感度不同，会对高注意力区域会分配更多的视觉处理资源，从而导致较高的视觉敏感度。门控卷积中的门控机制能根据输入特征动态地分配特征重要性，以模拟人眼视觉系统注意力机制导致的视觉敏感差异。门控卷积可以表示为:

$$F_o(x,y) = \sum_{i,j \in [1,k]} \phi(F_i(x+i, y+j) \cdot K_f(i,j)) \odot \sigma(F_i(x+i, y+j) \cdot K_g(i,j))$$

其中ϕ表示ReLU激活函数，σ表示sigmoid激活函数以控制门控值介于0和1之间。K_f,K_g是分别为特征卷积核和门控卷积核。

空间可分离卷积将卷积核分解为垂直和水平方向的两个独立卷积核，能够准确地捕捉图像中水平和垂直方向的特征。神经生理学的发现已经证实，垂直和水平方向的物体比倾斜方向的物体具有更高的视觉灵敏度。空间可分离卷积能更好地捕捉到图像在垂直和水平方向上的特征，与人眼的方向感知特性相契合。空间可分离卷积可以表示为

$$F_o(x,y) = \sum_{i \in [1,k]} \left(\sum_{j \in [1,k]} F_i(x+i, y+j) \cdot K_v(1,j) \right) \cdot K_h(i,1)$$

其中K_v,K_h分别表示垂直方向和水平方向的一维卷积核。

平均池化通过计算区域内的特征平均值，以平滑高频信息。相对于高频细节特征，人眼对低频的整体特征具有更高的分辨能力。平均值化能模糊高频特征，与人眼感知

特性相吻合。

最大池化通过选取区域内特征值的最大值，突出特征以聚焦显著特征。人眼对具有高视觉响应的显著特征具有更高的敏感度。最大池化通过选择具有最大响应的特征以突出显著特征，与人眼感知能力相一致。

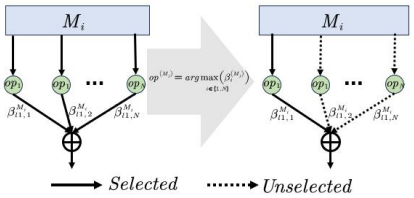


Figure 4 路径选择步骤的局部示例图。左边为搜索过程中的混合路径，右边表示最终选择的特征映射路径。

基于上述特征映射微元搜索空间，我们遵循DARTS的持续松弛方法为每种视觉掩蔽特征选择最优的微元。假设$O=\{o_i\}_{i\in[1,N]}$为上述的特征映射微元空间，每个特征映射微元记为o(·)。我们将映射微元的选择放宽为混合映射微元的优化。具体地，每种映射微元被添加一个可学习的权重，上述所有映射微元结果的softmax加权和即为混合映射微元。对于掩蔽特征M_i的混合映射微元下面的公式所示:

$$\overline{o}(M_i)=\sum_{o\in O}\frac{\exp\left(\beta_o^{(k)}\right)}{\sum_{o'\in O}\exp\left(\beta_{o'}^{(k)}\right)}o(M_i)$$

其中，视觉掩蔽特征M的混合映射微元权重由维度为$|\mathcal{O}|$的向量β参数化。在上述持续松弛方法下，混合映射微元优化变得可微分，从而可以在训练中通过梯度更新实现映射微元选择。具体地，在搜索训练过程中，连续变量β不断进行更新以调整表征微元的权重。在搜索训练完成后，混合映射微元会被最可能的映射微元替换，也即β中权重最大值所对应的表征微元。特征映射微元的选择过程可以公式化为:

$$o^*=arg\max_{o\in\mathcal{O}}\beta_o.$$

3.3.2、可扩展的掩蔽次序整理（Organizing）/排列（Arranging）

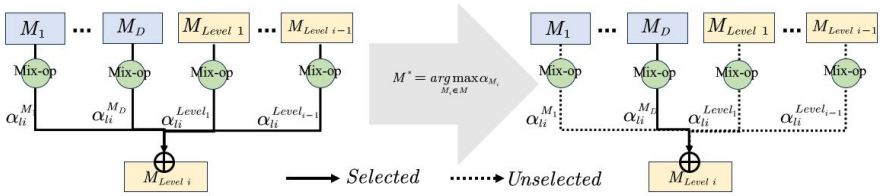


Figure 5 骨架搜索步骤的局部示例图。左边为搜索过程中的混合加权特征，右边表示最终选择的视觉掩蔽特征。

根据前面的讨论可知，掩蔽次序需要根据输入特征不同进行扩展和调整。为了实现可扩展的掩蔽次序整理，融合网络被定义为一个层级有序的融合结构，其中每层选择最合适的两种视觉掩蔽特征进行融合，从而确定掩蔽次序以形成一个完整的融合网络。该网络可以通过增加融合层的方式来融合任意数量的视觉掩蔽特征。更具体地说，对于 n 种视觉掩蔽特征融合任务，融合网络包含 n-1 层融合结构。这 n 种输入特征经过 n-1 融合层进行有序连接形成一个有向无环图，这个图即为融合骨架的搜索空间，如图 2 左所示。为了充分融合所有特征，我们允许之前的融合结果也作为候选特征，同时禁止重复选择同一特征。对于每层融合层，我们从输入特征和已融合结果中选择最合适的两种特征。第 n-1 层的融合输出即为所求的视觉冗余预测图。经过搜索，最终的融合网络是对上述有向无环图采样所得，最终会形成一个形如满二叉树的融合结构，如图 2 右所示。

我们在掩蔽次序整理中同样遵循 DARTS 的持续松弛方法。每个输入特征从包含给定的 n 种视觉掩蔽特征和前 i-1 层融合结果中选择。输入特征可以表示为 $V=\{M_1,M_2,M_3,...,M_n,M_{level\_1},M_{level\_2},...,M_{level\_i-1}\}$。与特征映射微元类似，我们将当前层的掩蔽次序整理放宽为所有输入特征的混合加权优化。具体地，每种输入特征被添加一个可学习的权重，所有输入特征的 softmax 加权和即为混合加权特征。可以由如下面的公式表示:

$$\overline{f}(M_i)=\sum_{M_i\in M}\frac{\exp\left(\alpha_{M_i}^{(k)}\right)}{\sum_{M_i'\in M}\exp\left(\alpha_{M_i'}^{(k)}\right)}\overline{o}(M_i),$$

其中，所有输入特征的权重由维数为$|M|$的向量α参数化。在上述持续松弛方法下，掩

蔽次序整理可以对混合特征中的权重进行梯度更新实现。具体地，在搜索训练过程中，连续变量α不断进行更新以调整每种输入特征的权重。在搜索完成后，最可能的视觉掩蔽特征会被选择在当前层进行融合。换句话说，在当前融合层中，连续变量α中最大的权重值所对应的视觉掩蔽特征被选择以确定掩蔽次序。每个骨架结构的搜索过程可以表示为：

$$M^* = \underset{M_i \in M}{arg\max} \alpha_{M_i}.$$

3.3.3、可通用的融合算子选择

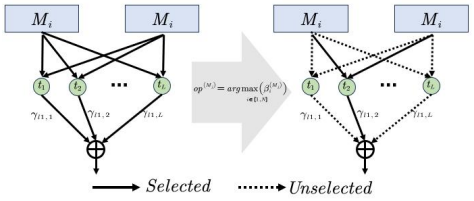


Figure 6 融合算子选择步骤的局部示例图。左边为搜索过程中的混合加权策略，右边表示最终选择的融合算子。

众所周知，视觉掩蔽特征在大多数图像中共存，如何有效的融合这些视觉掩蔽特征是获得准确视觉冗余的重要问题。因此，我们引入了融合算子选择这一步骤。融合算子选择的目标是为两种视觉掩蔽特征选择合适的融合运算。对于每种融合算子，输入包含两种视觉掩蔽特征，输出为两种掩蔽特征在所选择策略下的合成结果。为了保证我们融合算子选择的通用性，我们综合了现有融合方法，设计融合算子的搜索空间为$T = \{WeightedSum, ConcatFC, Attention, NAMM_\theta\}$，也即加权和，拼接，注意力，非线性加性融合，取最小值五种。每种融合算子都包含两个输入F_a, F_b。

加权和：现有融合方法中常用加权相加减的方法融合两种视觉掩蔽特征，这些方法可以总结为加权和的特殊情况，我们采用两个可学习参数w_1, w_2学习两种特征的融合权重。

$$WeightedSum = w_1 F_a + w_2 F_b.$$

拼接：拼接是深度视觉掩蔽特征融合常用的方法，拼接方法是将两种视觉掩蔽特征串联后经过一个全连接层，全连接层将串联后特征的通道数压缩到原始单一特征的特征数。

$$ConcatFC = FC(Concat(F_a, F_b)).$$

注意力：除了上述常用的视觉掩蔽特征融合方法外，我们还引入近期提出的注意力融合方法。由于标准的注意模块通常有三个输入，即查询、键和值，我们设查询为输入的一种特征，键和值为另一种特征。

$$Attention = Soft\max\left(\frac{F_a F_b^{\,T}}{\sqrt{C}} F_b\right).$$

其中 C 为通道数。除了上述融合算子，我们还采用了传统的 NAMM 模型。其中 NAMM 中的叠加因子在我们的方法中是可学习的参数。

与掩蔽次序类似，我们将当前层的融合算子选择放宽为在所有融合算子下输出的混合加权算子优化。具体地，每种融合算子被添加一个可学习的权重，所有融合算子输出结果的 softmax 加权和即为混合加权算子。混合加权算子可以由如下面的公式表示:

$$\overline{t}(M) = \sum_{t_i \in T} \frac{\exp\left(\gamma_{t_i}^{(k)}\right)}{\sum_{t_i' \in T} \exp\left(\gamma_{t_i'}^{(k)}\right)} t_l\left(\overline{f}(M_i), \overline{f}(M_j)\right),$$

其中，所有融合算子的权重由维数$|T|$的向量γ参数化。在上述持续松弛方法下，融合算子的选择可以对混合加权算子中的权重进行梯度更新实现。具体地，在搜索训练过程中，连续变量γ不断进行更新以调整每种融合算子的权重。在搜索完成后，最可能的算子会被选择用来融合输入的两种视觉掩蔽特征。换句话说，在当前融合层中，连续变量γ中最大的权重值所对应的融合算子被选择。融合算子选择可以表示为:

$$t^* = arg\max_{t \in T} \gamma_t.$$

经过上述映射微元选择，掩蔽次序整理和融合算子选择后，感知掩蔽融合网络基于超网络S_{hyper}和超参数α，β，γ推理得到。最终我们将训练集和验证集结合起来，对搜索到的融合网络进行再训练以实现视觉冗余预测。

4、实验评估和应用

4.1、实验配置

数据库描述：我们基于最新的基准数据集 SHEN2020 进行实验。该数据集涵盖了各种图像内容，包括室外、室内、景观、自然、人、物体和建筑。数据集包含 202 张大小为 1920×1080 的高清原始图像和 7878 张经过 VVC 压缩的图像组成。每个原始图像有 39

个不同冗余去除水平的编码版本，其中视觉无损编码版本通过主观实验选择。我们从测试集中选择15张代表性图片用于测试，这些图像如下图所示。

评估指标：为了评估视觉效应融合模型的性能，我们在预测的视觉阈值指导下，按照下列公式将噪声随机注入到每个像素中:

$$I_{con} = I_{ori} + r \times Sign \times I_{pre}$$

其中I_{rec}表示JND注入后的图像，I_{ori}表示原始图像。$Sign$表示与I_{ori}大小相同的随机矩阵，其值为1或-1。I_{pre}表示通过预测的可见性阈值。r表示噪声水平的调整因子。

对比方法：为了展示整体性能，我们将提出的方法与现有的xx种代表性方法进行了比较。考虑到不同方法对视觉掩蔽特征掩蔽存在差异，对比方法分为两类:

手工视觉掩蔽特征：Yang2005SPIC, Liu2010TCSVT, Wu2013TMM, Wu2013TIP, Wang2016TIP, Wu2017TIP, Chen2020TCSVT, Wang2022TII, Huang2023ICME。每种方法所采用的视觉掩蔽特征如下表所示。

深度视觉掩蔽特征：Shen2021TIP。

为了公平起见，对于所有对比方法，我们使用相关作者开源的代码提取了对应的视觉特性，不做任何额外的调整。

Method	Masking Effects
Yang2005SPIC	The LA and CM effects
Liu2010TCSVT	The LA, EM, and TM effects
Wu2013TMM	The LA, CM, and DC effects
Wu2013TIP	The LA and PM effects
Wang2016TIP	The LA, CM, and SS effects
Wu2017TIP	The LA, CM, and PM effects
Chen2019TCSVT	The LA and CM with asymmetry effects
Wang2021SPL	The LA, CM, OC, and HFS effects
Wang2022TII	The LA, CM, OC, and BS effects
Huang2023ICME	The LA, CM of certainty and uncertainty, OC and BS effects

实现细节：所有的实验都是在 Intel(R) Xeon(R) Gold 6226R CPU@2.90GHz, 64GB RAM, Nvidia Tesla A100 GPU 上进行的。我们使用 Python 工具箱 PyTorch 来训练所有的模型。在训练过程中，图像样本的大小为原始大小。所有权重都由截断正态初始化器初始化，Adam 优化器使用默认参数（β1=0.9 和β2=0.999）。整个训练过程分为搜索阶段和再训练阶段。

搜索阶段： 我们首先在 SHEN2020 数据库上进行融合架构搜索。在架构搜索阶段中，我们将数据集以 5:5 的比例划分成训练集和验证集，设置 batchsize 为 4，进行 30 个 epoch 的搜索。我们将初始学习率设置为 0.025 并采用余弦衰减随着训练到 0，每个网络搜索需要大约 2 个小时。

再训练阶段： 与 DARTS 相同，我们对搜索阶段获得的融合网络结构进行再训练以微调参数和权重。再训练阶段中，我们将数据集按照 8:1:1 的比例划分成训练集，验证集和测试集，设置 Batchsize 为 8，并对我们的模型进行 100 个 epoch 的再训练。我们将初始学习率设置为 0.01 采用余弦衰减随着训练衰减到 0，每个网络再训练需要大约 1 个小时。

4.2、整体性能对比

定量指标对比：在相同噪声水平下，更好的 JND 模型能引导更多的噪声注入到视觉冗余区域，从而具有更高的感知质量。我们进行了相同噪声水平下（MSE=100）的对比实验，噪声按照公式 1 注入到图像中。

为了评估噪声注入后的感知质量，我们进行了一个主观实验，主要依据 ITU-R BT.500-13 标准进行。我们邀请了 20 名受试者参加主观测试，其中 12 名男性和 8 名女性，年龄从 18 岁到 40 岁。两张图像并列显示在两个 27 寸的 Samsung C27F390FHC 监视器上，其中一张是使用原始融合方法得到的 JND 污染的图像，另一张是使用提出方法融合得到的 JND 污染的图像。其他主观实验设置与 SHEN2020 中相同。在获得主观结果后，我们使用 SHEN2020 中建议的基于统计的异常值检测方法来检查实验中原始结果的一致性。在去除异常数据后，平均结果被视为最终 MOS 分数。

每个模型的平均 MOS 结果如下表所示，其中正（负）值表示提出的模型能较高

（差）。可以发现我们的NAS-NAMM相较于十种原始JND模型都有更高的平均MOS。

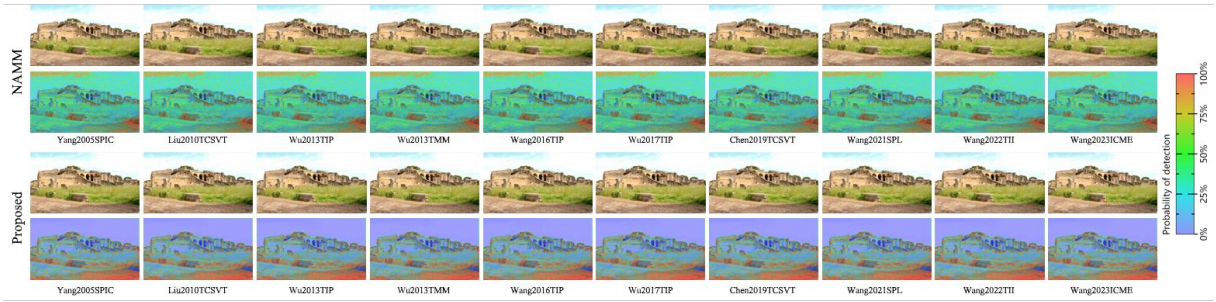
Method	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
Yang2005SPIC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Liu2010TCSVT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wu2013TMM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wu2013TIP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wang2016TIP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wu2017TIP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Chen2019TCSVT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wang2021SPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wang2022TII	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Huang2023ICME	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

不失公平性地，我们还提供了在常用客观评估指标VMAF的性能对比，结果如下表所示。从表中可以看出，我们方法在客观指标下仍然具有更好的性能。

Method		I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
Yang2005SPIC	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Liu2010TCSVT	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wu2013TMM	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wu2013TIP	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wang2016TIP	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wu2017TIP	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Chen2019TCSVT	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wang2021SPL	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Wang2022TII	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
Huang2023ICME	NAMM	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000
	Proposed	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000	00.0000

定性质量对比：为了更直观的对比JND模型性能，我们在下图中提供了每种对比方法在原始融合方法和提出融合搜索方法下的视觉冗余预测图，其中像素越亮意味着冗余度越高。值得注意的是，现有方法容易在显著物体或平坦区域预测较高的阈值，同时低估高复杂度区域的视觉阈值。而平滑区域微小的变化都可能引起HVS的高度关注，高估的视觉阈值导致重建图像视觉质量变差。另一方面，HVS难以察觉纹理区域的噪声，低估的视觉阈值导致大量视觉冗余残留。从下图中可以看出，由于更优的视觉掩蔽融合方法，提出的方法能够合理的抑制平滑区域的视觉阈值以保证视觉质量，同时增高纹理区域的视觉阈值以充分去除视觉冗余。

为了更清楚地对比我们的方法与对应原始方法，下图中展示了我们的方法和原始方法在 HDR-VDP2.2 下的失真可感知概率。从下图中可以看出，在注入等量噪声条件下，提出的方法中失真被检测到的概率更小。主要原因可以解释为我们的融合方法倾向于将噪声引入建筑物，草地等纹理区域，这和人眼难以感知纹理区域的噪声相符。



4.3、消融实验

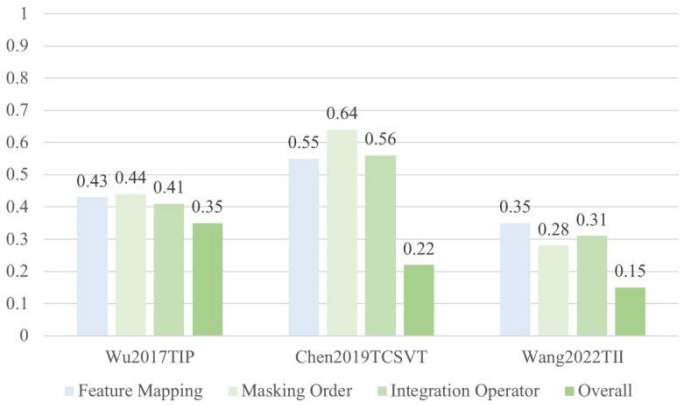
为了验证引入的网络层搜索空间的必要性和进行架构搜索的有效性，我们进行了两个额外的消融实验以研究我们的融合方法。我们计算预测 JND 图与标签 JND 图之间的 RMSE 值。RMSE 值越小，意味着预测的 JND 图与标签 JND 图的相似度越高，从而直接表明特定模型预测 JND 的能力越好

特征映射微元的作用：为了验证我们设计的网络层搜索空间的必要性，我们消融掉网络层搜索空间进行了额外的验证实验，实验结果如下图所示。具体地，我们采用随机抽样的方作为映射微元选择的替代方案，按照 4.1 节中的描述进行训练，并对比了测试结果与标签之间的 RMSE。实验结果表明，经过映射微元选择预测得到 JND 图与标签具有更高的一致性，这验证了所提出的网络层搜索空间的必要性。

掩蔽次序整理的作用：为了验证我们提出的掩蔽次序整理的有效性，我们消融掉架构搜索进行了额外的验证实验，实验结果如下图所示。具体地，我们采用随机抽样的方

式生成掩蔽融合次序，并按照 A 节中的描述进行再训练，并对比了测试结果与标签之间的 RMSE。实验结果表明，经过掩蔽次序整理得到的网络经过训练后预测的 JND 图与标签分布更加相似，这验证了所提出的融合架构搜索的有效性。

融合算子选择的作用：为了验证我们方法中融合算子选择的重要性，我们消融掉融合算子选择进行了额外的验证实验，实验结果如下图所示。具体地，我们采用随机抽样的方作为融合算泽选择的替代方案，按照 4.1 节中的描述进行训练，并对比了测试结果与标签之间的 RMSE。实验结果表明，经过融合算子选择得到 JND 图与标签具有更高的一致性，这验证了所提出的网络层搜索空间的必要性。



4.4、压缩应用

JND 图提供了人眼视觉可见度限制的信息，因此经常被用来指导压缩过程。在本节中，我们将提出的 NAS-NAMM 融合方法和 NAMM 融合方法得到的 JND 配置文件合并进广泛使用的 JPEG 和两种最先进的压缩标准（HEVC 和 VVC）。

JND 引导的 JPEG 压缩：在 JND 引导的 JPEG 压缩中，一种广泛使用的方式是在编码前去除输入图像的视觉冗余，以节省比特率，同时保持相同的视觉质量。具体地，每个 8×8 块由预测的可见性阈值处理如下：

$$\widehat{I}_{ori}(\mathbf{p}) = \begin{cases} \overline{I}_b, & if\ |I_{ori}(\mathbf{p}) - \overline{I}_b| \le I_{vt}(\mathbf{p}), \\ I_{ori}(\mathbf{p}) + I_{vt}(\mathbf{p}), & if\ I_{ori}(\mathbf{p}) - \overline{I}_b < -I_{vt}(\mathbf{p}), \\ I_{ori}(\mathbf{p}) - I_{vt}(\mathbf{p}), & otherwise. \end{cases},$$

其中 $I_{ori}(p)$ 表示 p 位置的像素大小，$I_{vt}(p)$ 表示相关的视觉阈值，I_b 表示当前 8×8 块的平均亮

度值，$\hat{I}_{ori}(p)$表示预处理后的输出。

JND 引导的 HEVC/VVC 压缩：在 HEVC 或 VVC 标准中，我们首先获得 JND 引导的残差值，然后用 HEVC 或 VVC 编码器对残差块进行编码。具体地，我们计算 JND 引导的残值如下列公式所示:

$$\widehat{R}(\mathbf{p}) = \begin{cases} 0, & if\ |R(\mathbf{p})| \le I_{vt}(\mathbf{p})\ and\ \sigma^2(\mathbf{p}) > \sigma^2, \\ \min(R(\mathbf{p}) \times \dfrac{|I_{vt}(\mathbf{p})|}{\sigma^2},\ R(\mathbf{p}) + I_{vt}(\mathbf{p})),\ else\ if\ R(\mathbf{p}) < 0, \\ \max(R(\mathbf{p}) \times \dfrac{|I_{vt}(\mathbf{p})|}{\sigma^2},\ R(\mathbf{p}) - I_{vt}(\mathbf{p})),\ otherwise. \end{cases},$$

其中，$\delta^2(p)$、δ^2表示局部范围与 p 位置编码块的相关方差，$R(p)$表示 HEVC 或 VVC 得到的残差，$\hat{R}(p)$表示处理后的预测残差。

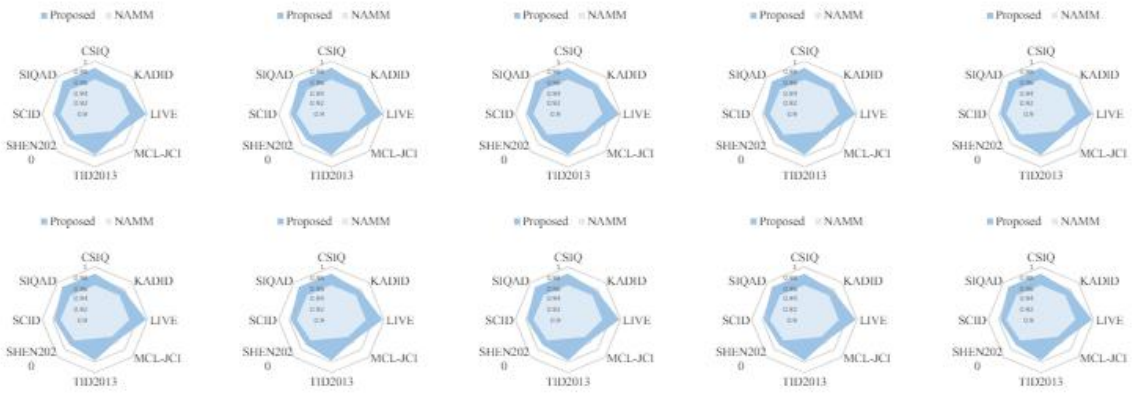
下图中显示了十种方法分别在原始方法和本文提出的融合方法下引导压缩的视觉性能比较。很明显，提出的方法引导的压缩提供了几乎相同的视觉质量的无损编码结果，平均节省了 21.32%的比特率，相较于原始方法更高。

4.5、跨库验证

这项工作的目的是为开发一个通用性的视觉特性融合方案。为了验证提出融合方法的泛化性，我们进行了交叉数据集评估，在以下 8 个基准数据集上进行测试：1）CSIQ（30 张 512×512 大小的原始自然内容图像），2）KADID-10K（81 张 512×384 大小的原始天然内容图像），3）LIVE（29 张原始自然内容各种分辨率的图像），4）MCL-JCI（50 张 1920×1080 大小的原始天然内容图像），5）TID2013（25 张 512×384 大小的原始自然内容图像），6）SHEN2020（测试集中 20 幅 1920×1080 大小的原始自

然内容图像），7）SCID（40 张 1280×720 大小的原始屏幕内容图像），8）SIQAD（20 张各种分辨率的原始屏幕内容图像）。

我们设置注入噪声水平为 MSE=100，并根据数据库类型分别采用相应的质量评估指标。具体的，对于自然图像数据库，我们采用 Mittal 等人基于自然场景统计提出的质量评估指标 BRISQUE；对于屏幕内容图像，我们采用 Huang 等人提出的屏幕内容图像质量评估指标 GBI，实验结果如下图所示。从下图中可以推断，在所有类型的基准数据集上，NAS-NAMM 融合方法仍然取得了令人满意的性能。主要原因可以解释为我们的融合网络架构主要学习到了各类视觉特性的更优融合结构，这意味着我们方法中的架构搜索方法和网络层搜索空间可以用于各种图像类型。



5、结论

本文提出了一种用于视觉掩蔽特征融合的通用架构搜索方法。具体而言，我们考虑 HVS 的感知特点，设计了多种可解释的特征映射微元将异质的视觉掩蔽特征映射到相似的空间以促进融合。为了有效地融合掩蔽特征，一个可扩展的掩蔽次序整理方法被提出。此外，我们设计了一个可通用的融合算子选择方法，以针对视觉掩蔽特征特点选择融合运算。与多种方法在 8 个基准数据集上的实验结果证明了该方法在定性和定量上的优越性。我们认为，提出的视觉掩蔽特征融合搜索方法将有助于提高图像压缩和处理系统的服务质量。

6、参考文献

[1]. Yi-Jen Chiu, Toby Berger. A software-only video codec using pixelwise conditional

differential replenishment and perceptual enhancements[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(3): 438-450.

[2]. Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11): 1648-1652.

[3]. Sung-Ho Bae, Munchurl Kim. A novel DCT-based JND model for luminance adaptation effect in DCT frequency. IEEE Signal Processing Letters, 2013, 20(9): 893-896.

[4]. Sung-Ho Bae, Munchurl Kim. A new DCT-based JND model of monochrome images for contrast masking effects with texture complexity and frequency. IEEE International Conference on Image Processing. IEEE, 2013, 1(1): 431-434.

[5]. Wei, Zhenyu, and King N. Ngan. Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(3): 337-346.

[6]. Hao Chen, Ruimin Hu, Jinhui Hu, Zhongyuan Wang . Temporal color just noticeable distortion model and its application for video coding. IEEE International Conference on Multimedia and Expo (ICME). 2010: 713-718.

[7]. Sung-Ho Bae, Munchurl Kim. A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images. IEEE Transactions on Image processing, 2014, 23(8): 3227-3240.

[8]. Miaohui Wang, Zhuowei Xu, Xueqin Liu, Jian Xiong, Wuyuan Xie. Perceptually quasi-lossless compression of screen content data via visibility modeling and deep forecasting. IEEE Transactions on Industrial Informatics, 2022, 18(10): 6865-6875.

[9]. Chou C H, Li Y C. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. IEEE Transactions on circuits and systems for video technology, 1995, 5(6): 467-476.

[10]. Jinjian Wu, Leida Li, Weisheng Dong, Guangming Shi, Weisi Lin, C.-C. J. Kuo. Enhanced just noticeable difference model for images with pattern complexity. IEEE Transactions on Image Processing, 2017, 26(6). pp. 2682-2693.

[11]. Heidi Peterson, Albert Ahumada, Andrew Watson Heidi. Improved detection model for DCT coefficient quantization. Human Vision, Visual Processing, and Digital Display IV. SPIE, 1993, 1913: 191-201.

[12]. Yang X K, Ling W S, Lu Z K, et al. Just noticeable distortion model and its applications

in video coding. Signal processing: Image communication, 2005, 20(7). pp. 662-680.

[13]. Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu. Just noticeable difference estimation for images with free-energy principle. IEEE Transactions on Multimedia, 2013, 15(7): 1705-1710.

[14]. Miaohui Wang, Xueqin Liu, Wuyuan Xie, Long Xu. Perceptual redundancy estimation of screen images via multi-domain sensitivities. IEEE Signal Processing Letters, 2021, 28. pp. 1440-1444.

[15]. Shiqi Wang, Lin Ma, Yuming Fang, Weisi Lin, Siwei Ma, Wen Gao . Just noticeable difference estimation for screen content images. IEEE Transactions on Image Processing, 2016, 25(8). pp. 3838-3851.

[16]. Hongkui Wang, Li Yu, Junhui Liang, Haibing Yin, Tiansong Li , and Shengwei Wang. Hierarchical predictive coding-based JND estimation for image compression. IEEE Transactions on Image Processing, 2020, 30: 487-500.

[17]. Wuyuan Xie, ShukangWang, Sukun Tian, Lirong Huang, Ye Liu, Miaohui Wang. Just Noticeable Visual Redundancy Forecasting: A Deep Multimodal-driven Approach. arXiv preprint arXiv:2303.10372, 2023.

[18]. Chunling Fan, Hanhe Lin, Vlad Hosu, Yun Zhang, Qingshan Jiang, Raouf Hamzaoui, and Dietmar Saupe. SUR-Net: Predicting the satisfied user ratio curve for image compression with deep learning. IEEE eleventh international conference on quality of multimedia experience (QoMEX). 2019: 1-6.

[19]. Huanhua Liu, Yun Zhang, Huan Zhang, Huan Zhang, Sam Kwong, C.-C. Jay Kuo, Xiaoping Fan. Deep learning-based picture-wise just noticeable distortion prediction model for image compression[J]. IEEE Transactions on Image Processing, 2019, 29: 641-656.

[19]. Yuhao Wu, Weiping Ji and Jinjian Wu. Unsupervised deep learning for just noticeable difference estimation. IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2020: 1-6.

[20]. Jian Jin,, Dong Yu, Weisi Lin, Lili Meng, Hao Wang, Huaxiang Zhang. Full RGB just noticeable difference (JND) modelling. arXiv preprint arXiv:2203.00629, 2022.

[21]. Sanaz Nami, Farhad Pakdaman, Mahmoud Reza Hashemi, Senior Member, IEEE, Shervin Shirmohammadi. BL-JUNIPER: A CNN-assisted framework for perceptual video coding leveraging block-level JND. IEEE Transactions on Multimedia, 2022.

[22]. Xuelin Shen, Zhangkai Ni, Wenhan Yang, Xinfeng Zhang, Shiqi Wang, Sam Kwong.

Just noticeable distortion profile inference: A patch-level structural visibility learning approach. IEEE Transactions on Image Processing, 2020, 30: 26-38.