arXiv:2301.01182v2 [eess.IV] 3 Nov 2023

PMT-IQA：用于盲图像质量评估的渐进式多任务学习

Qingyi Pan[1,2], Ning Guo[1], Letu Qingge[3], Jingyi Zhang[2], 和 Pei Yang[1](✉)

[1] 青海大学计算机技术与应用系，中国[2] 清华大学统计科学中心与工业工程系，中国[3] 北卡罗来纳州立农业技术大学计算机科学系，美国

摘要。 由于失真类型多样且图像内容各异，盲图像质量评价（BIQA）仍面临挑战。这使得跨越不同尺度的失真模式更加复杂，并加剧了 BIQA 回归问题的难度。然而，现有 BIQA 方法往往未能考虑多尺度失真模式和图像内容，关于通过特定学习策略提升质量回归模型性能的研究也十分有限。本文提出一种简单而有效的渐进式多任务图像质量评价（PMT-IQA）模型，该模型包含多尺度特征提取模块（MS）和渐进式多任务学习模块（PMT），旨在帮助模型学习复杂的失真模式，并更好地优化回归问题，使其符合人类从易到难的学习规律。为验证所提出的 PMT-IQA 模型的有效性，我们在四个广泛使用的公共数据集上进行了实验。实验结果表明，PMT-IQA 的性能优于对比方法，且 MS 和 PMT 模块均能提升模型性能。本研究的源代码见 https://github.com/pqy000/PMT-IQA。

关键词： 盲图像质量评价 · 由易到难效应 · 多尺度特征 · 渐进式多任务学习。

1 引言

近年来，随着智能手机和其他相机设备的普及，大量图像被生成，并在人们的信息交互中发挥着日益重要的作用。然而，这些图像可能会受到多种因素的影响而产生失真（即由噪声、有损压缩等造成的质量下降），包括摄影师的专业水平、设备性能、传输和设备存储等。因此，评估图像质量十分必要。尽管人们能够准确可靠地对图像质量进行主观评价，但由于耗时费力 [28]，这种方法在实际应用中受到很大限制。因此，客观图像质量评价（IQA） [9], 旨在探索能够自动评估图像质量的模型

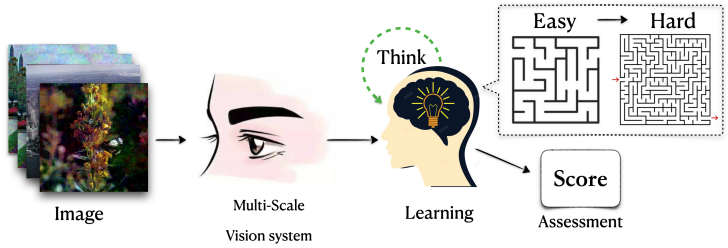


图1. 所提出的 PMT-IQA 的动机示意图。我们将图像质量评估分为两部分：多尺度视觉系统和基于人类学习规律的图像质量评估。

，使其评价结果符合人类视觉系统（HVS）的感知，这在过去几年中引起了广泛关注 [28] [25]。在所有客观图像质量评价方法中，盲图像质量评估（BIQA），又称无参考图像质量评估（NR-IQA），是最具挑战性的一类。“盲”在此处指图像质量评估过程中不需要无失真图像（即未失真的参考图像）。尽管这一领域已取得很大进展，但它仍然是一个开放且富有挑战性的问题；本研究致力于探索盲图像质量评估问题。

失真的多样性和图像内容的可变性是盲图像质量评估面临诸多挑战的主要原因。一方面，它们使失真模式变得复杂，涵盖从局部到全局的多个尺度。另一方面，复杂的输入空间加剧了盲图像质量评估中回归问题的难度。然而，现有研究往往未能考虑多尺度失真模式和图像内容。已有一些工作尝试为图像质量评价设计端到端架构。例如，Li 等人 [15] 使用预训练的深度卷积神经网络（DCNN）提取特征。显然，随着网络加深，卷积层的感受野逐渐扩大，因此 DCNN 能够学习全局特征。然而，大多数真实世界图像数据中的失真模式都存在于局部区域。因此，仅凭全局特征不足以捕捉复杂失真。此外，人类学习过程遵循由易到难的规律，这被称为“由易到难效应”，由 Pavlov 于 [19]1927年提出。然而，现有的盲图像质量评估方法往往直接解决复杂的回归问题。

本文受人类视觉系统（HVS）的多尺度特性以及图 1 所示人类由易到难的学习规律启发，提出了一种简单而有效的图像质量评价架构。我们将所提出的网络命名为渐进式多任务图像质量评估（PMT-IQA），因为该网络旨在通过模拟人类由易到难学习规律的任务迁移策略，捕捉与失真相关的模式。所提出模型的思路如下。首先，我们设计了一个多尺度语义特征提取模块，以提取由全局到局部的失真感知特征。其次，受由易到难的学习规律启发，我们构建了一种渐进式多任务学习方案，

该方案可以逐步从简单任务（即质量等级分类）过渡到困难任务（即质量分数回归）。最后，我们在多个广泛使用的公开图像质量评价数据集上评估了所提出的 PMT-IQA 的性能，实验结果验证了 PMT-IQA 模型的有效性。

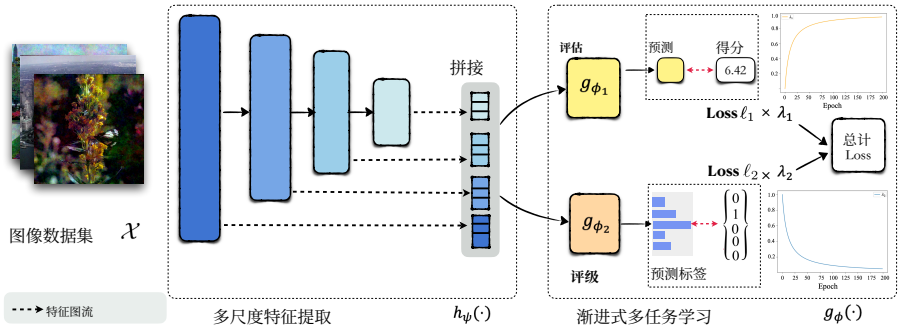


图 2. 渐进式多任务学习图像质量评价架构。该架构将图像质量评价任务分为两个步骤：多尺度语义特征提取和渐进式多任务学习。

2 相关工作

近年来，深度神经网络技术的发展极大地推动了盲图像质量评估问题的研究。Zhang 等人 [30]提出了一种双分支网络，可同时适应真实失真和合成失真。Su 等人 [22] 针对真实场景失真图像提出了一种自适应超网络，以应对真实世界图像中失真类型多样、内容各异的挑战。为解决盲图像质量评估模型跨场景能力较弱的问题，Zhang 等人 [31]提出了一种统一的不确定性感知盲图像质量评价模型，适用于实验室和自然场景。此外，一些最新的深度学习技术，包括对比学习 [17], 图卷积神经网络 [23], 和 Transformer [5], 也已应用于盲图像质量评估研究。尽管这些模型取得了令人瞩目的性能，但它们主要关注网络架构设计，很少研究如何使模型学习得更好。此外，这些现有模型也很少使用多尺度特征提取。

受由易到难效应的启发 [19], ，我们提出了一种新的学习方案，通过模拟由易到难的学习规律来进行盲图像质量评估。事实上，课程学习 [2], 在过去十年中受到了广泛关注 [24],，它正是这一理念的一种实现。需要注意的是，课程学习通过训练数据来模拟由易到难的学习规律，即

先用较容易的样本训练机器学习模型，再逐步使用较难的样本。然而，正如由易到难效应所指出的，“先通过较容易的版本体验一项辨别任务，有助于随后学习难度更高的任务 *[19]*”，我们的想法是通过渐进式学习不同难度级别的相关任务，来模拟由易到难的学习规律。我们在实现中采用了多任务架构。不同于现有基于多任务的盲图像质量评估模型 [7,16], 同时预测失真类型和图像质量分数，我们模型中的两个任务（即质量等级预测和质量分数预测）是渐进式的。另一方面，我们采用动态任务权重而非固定权重，实现从质量等级分类到质量分数回归的过渡。总之，我们提出了一种新颖的渐进式多任务学习方案，通过模拟由易到难的学习规律来进行盲图像质量评估。

3 方法

3.1 所提出模型概述

所提出的渐进式多任务图像质量评价（PMT-IQA）模型架构如图 2 所示。它包含一个多尺度特征提取模块（MS）和一个渐进式多任务学习模块（PMT）。MS 模块旨在利用多尺度信息提取表征能力更强的图像特征，而 PMT 模块则通过动态加权的双头多任务结构来模拟由易到难的学习规律。所提出的 PMT-IQA 模型可用数学形式表示，如式（1）所示。

$$f_\theta(\cdot) = g_\phi \circ h_\psi(\cdot) \tag{1}$$

其中，$f_\theta(\cdot)$ 表示参数为 θ 的完整模型，h_ψ 是 MS 模块，可获取由局部到全局的失真；PMT 模块 g_ϕ 用于学习复杂的回归问题。式（1）中的算子 $\circ$ 表示模块 g_ϕ 和 h_ψ 的组合，其中 h_ψ 的输出作为 g_ϕ 的输入。参数 $\theta = \{\phi, \psi\}$ 的定义将在下一节中说明。

3.2 多尺度语义特征提取

为了表征各种失真，我们利用卷积提取多尺度特征（从局部到全局），每个尺度都对应一个特征图 s_i。假设我们有n 个尺度的特征，那么我们将所有特征拼接起来，如公式（2）所示。

$$h_\psi(x_i) = \mathrm{concat}(s_1, \cdots s_j, \cdots, s_n) \tag{2}$$

更具体地说，我们在 PMT-IQA 中使用预训练的 ResNet50 [6]作为骨干架构，并收集 ResNet50 四个阶段的特征图。随后，我们使用 1×1 卷积和全局平均池化进行维度对齐。MS 模块 $h(\cdot)$的输出被送入 PMT 模块进行预测。

3.3 渐进式多任务图像质量评价

如第 1 节所述，失真的多样性和图像内容的可变性使质量标量得分回归问题的输入空间变得复杂，并增加了模型学习的难度。为应对这一挑战，我们提出了一种渐进式多任务学习方案，以模拟人类从易到难的学习过程。具体而言，除了质量分数预测任务，我们还引入了一个与质量回归相关且相对简单的质量等级分类任务。在模型训练过程中，我们动态调整质量分数预测任务和质量等级分类任务在学习目标函数中的权重，旨在逐步将模型的注意力从质量等级分类任务转移到质量分数预测任务。

对于质量等级分类任务，我们将标量质量分数的取值范围划分为若干离散子区间，并将每个子区间视为一个质量类别，用以表示质量分类任务中的特定质量等级。设w 为区间长度， $[y^{min}, y^{max}]$ 为质量分数的取值范围，则可得到 K 个类别，如下所示：

$$K = \lfloor \frac{|y^{\max} - y^{\min}|}{w} \rfloor \tag{3}$$

对于标量质量分数为 y_i的样本 x_i ，我们可以通过将 y_i 映射到相应的离散质量区间，得到对应的质量类别标签 $y_i^c \in Y = \{1, \cdots, K\}$。

如图2所示，PMT g_ϕ 包含两个部分：标量图像质量分数评估模块 g_{ϕ_1} ：$\mathbb{R}^h \to \mathbb{R}$ ，以及图像质量等级分类模块g_{ϕ_2} ： $\mathbb{R}^h \to [0,1]^K$。g_{ϕ_1} 和g_{ϕ_2} 均采用简单的多层感知机（MLP）实现，其中 g_{ϕ_1} 由四个全连接层组成， g_{ϕ_2} 包含三个全连接层和一个Softmax层。ReLU()分别用作g_{ϕ_1} 和g_{ϕ_2}前3个和前2个全连接层的激活函数。假设 $\phi_1 = \{W_{(\phi_1)}\,1,\ W^{(\phi_1)}\,2,\ W^{(\phi_1)}\,3,\ W_{(\phi_1)\,4}\ \}$ 以及 $\phi_2 = \{W_{(\phi_2)}\,1,\ W^{(\phi_2)}\,2,\ W_{(\phi_2)\,3}\ \}$，其中 $W_{(\phi_1)\,i}$和 $W_{(\phi_2)\,i}$分别是 g_{ϕ_1} 和 g_{ϕ_2} 第i层的参数，则对于输入 X （注意， X 实际上是 [ˆX； 1] 对应于真实输入ˆX，因为 $W_i^{\phi_j}$ 表示权重和偏置）， g_{ϕ_1} 和 g_{ϕ_2} 定义如下：

$$g_{\phi_1}(X) = W_4^{(\phi_1)}(\mathrm{ReLU}(W_3^{(\phi_1)}(\mathrm{ReLU}(W_2^{(\phi_1)}(\mathrm{ReLU}(W_1^{(\phi_1)}X)))))) \tag{4}$$

$$g_{\phi_2}(X) = (\frac{\exp(o_1)}{\sum_{i=1}^K \exp(o_i)}, \cdots, \frac{\exp(o_K)}{\sum_{i=1}^K \exp(o_i)}) \tag{5}$$

其中， o_i 是g_{ϕ_2}最后一个全连接层输出的第i个分量，其定义如下：

$$(o_1, \cdots, o_K) = W_3^{(\phi_2)}(\mathrm{ReLU}(W_2^{(\phi_2)}(\mathrm{ReLU}(W_1^{(\phi_2)}X)))) \tag{6}$$

根据 g_{ϕ_1} 和 g_{ϕ_2}的定义，PMT-IQA 中的目标损失函数可定义为公式（7）。

$$\lambda_1 \mathcal{L}_r(x, y) + \lambda_2 \mathcal{L}_c(x, y) \tag{7}$$

其中， $\mathcal{L}_r$ 和 $\mathcal{L}_c$ 分别是图像质量分数回归任务和图像质量等级分类任务的损失项。参数 λ_1， $\lambda_2 > 0$ 是训练过程中的动态超参数。在我们的实现中，分别使用 ℓ_1 损失（定义见公式 (8)）进行 $\mathcal{L}_r$ ，并使用交叉熵损失（定义见公式 (9)）进行 $\mathcal{L}_c$。

$$\mathcal{L}_r(x, y) = \frac{1}{n} \sum_{i=1}^{n} |y_i - g_{\phi_1}(h(x_i))| \tag{8}$$

$$\mathcal{L}_c(x, y) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{K} (y_i)_c \log((g_{\phi_2}(h(x_i)))_c) \tag{9}$$

其中，公式 (9) 中的 $(\cdot)_c$ 表示 $(\cdot)$ 的第 c 个分量。

为模拟由易到难的学习规律 [19]，我们使模型在训练初期专注于学习分类任务，并随着训练的推进，通过动态调整分类和回归任务的权重，逐渐将重心转向标量质量分数评估，具体如下：

$$\begin{aligned} \lambda_1(t) &= \frac{t}{T+1}\xi \\ \lambda_2(t) &= 1 - \lambda_1(t) \end{aligned} \tag{10}$$

其中， t 表示第 t 个训练轮次， T 表示最大训练轮次。 ξ 是用于平衡两种损失尺度差异的权衡系数。我们采用 Adam 优化器 [10] ，联合优化 PMT-IQA 参数 ϕ 和 ψ 。

4 实验

4.1 实验设置

数据集 我们使用四个公开可用的图像质量评价数据集，包括 LIVE Challenge (LIVE-C) [4], BID [3],LIVE [20], 和 CSIQ [13]，来评估各图像质量评估算法。在这四个数据集中，BID 和 LIVE-C 是真实失真数据集，其中 BID 包含 586 张具有真实模糊失真的图像，LIVE-C 则包含 1162 张由不同相机拍摄的真实世界图像。除真实失真数据集外，我们还在两个合成图像数据集 LIVE 和 CSIQ 上评估 PMT-IQA，它们分别包含 779 张和 866 张图像，且分别具有 5 种和 6 种单一失真。

评估指标 我们选用两个常用的评估指标——斯皮尔曼等级相关系数（SRCC）[14]和皮尔逊线性相关系数（PLCC）[14], ，来评估图像质量评估算法的性能。SRCC 和 PLCC 的定义分别见公式 11 和公式 12。SRCC 和 PLCC 的取值范围均为 -1 到 1，数值越大表示性能越好。

$$\mathrm{SRCC} = 1 - \frac{6\sum_i d_i^2}{n\left(n^2-1\right)} \tag{11}$$

$$\mathrm{PLCC} = \frac{\sum_i \left(q_i - q_m\right)\left(\hat{q}_i - \hat{q}_m\right)}{\sqrt{\sum_i \left(q_i - q_m\right)^2 \sum_i \left(\hat{q}_i - \hat{q}_m\right)^2}} \tag{12}$$

其中， d_i 表示 MOS 与第i张图像的预测得分之间的秩差， n 为图像数量。 q_i 和$\hat{}q_i$ 分别表示第i张图像的 MOS（差分平均意见得分）和预测得分， q_m 和$\hat{}q_m$ 分别表示所有图像的平均 MOS（差分平均意见得分）值和平均预测得分。

实现细节 为进行公平比较，我们遵循 HyperIQA [22] 中的实验方案和设置。每个数据集均按 4:1 的比例划分为训练集和测试集。我们将质量得分缩放至 [0,1] 以提高稳定性，如图 3 所示。训练期间，按照 HyoerIQA [22]的方法，通过随机裁剪和水平翻转对每张训练图像进行数据增强：对 LIVE-C 数据集增强十次，对其他三个数据集增强五次。我们采用近期提出的超参数搜索框架 Optuna [1] 优化超参数，并在表 1 中报告了 PMT-IQA 在四个数据集上的超参数取值。此外，我们使用 Dropout [21] 和权重衰减 [12]策略来避免过拟合。

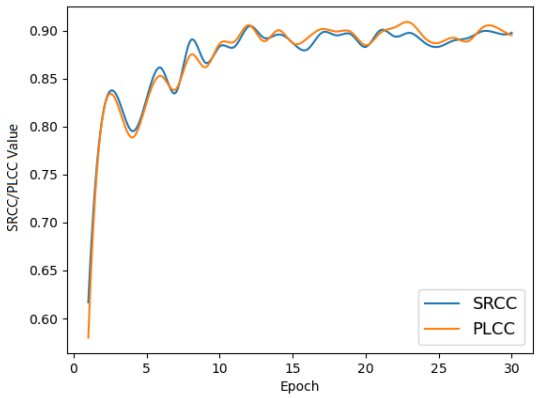


图 3. PMT-IQA 在 BID 数据集训练过程中的 SRCC 和 PLCC 值。

表 1. Optuna 在四个测试数据集上获得的超参数。

数据集	LR	批次大小	ξ	优化器
BID	1.09e-4	12	0.9419	Adam
LIVE-C	4.72e-4	12	0.9841	Adam
LIVE	3.23e-4	12	0.9941	Adam
CSIQ	4.72e-4	12	0.8931	Adam

4.2 性能评估

我们选取了十四种具有代表性的盲图像质量评估方法作为强基线方法，包括 BRISQUE [18], 、ILNIQUE [29],、AlexNet [11],、ResNet50 [6], 、HOSA [26],、BIECON [8], 、SFA [15], 、PQR [27],、DB-CNN [30], 、HyperIQA [22], 、UNIQUE [31],、CONTRIQUE [17],、GraphIQA [23],和 TRes [5],，以评估我们所提出的 PMT-IQA 模型的性能。

表 2 列出了各方法在四个测试数据集上的 SRCC 和 PLCC 值。从表 2 可以看出，在 BID 和 LIVE-C 数据集上，无论是 SRCC 还是 PLCC 评价，PMT-IQA 方法的表现均优于所有对比方法。我们的 PMT-IQA 模型在 LIVE 数据集上取得了第二好的结果，仅次于 GraphIQA。对于另一个合成失真数据集 CSIQ，PMT-IQA 也取得了最佳 SRCC 值（0.949）和具有竞争力的 PLCC 值（0.951）。为了更直观地比较各方法的性能，我们还列出了所有方法的排名（即表 2 中括号内的数字）以及各方法的 SRCC 和 PLCC 平均排名（即表 2 的最后两列）。所提出的 PMT-IQA 在四个测试数据集上的 SRCC 和 PLCC 平均排名分别为 1.25 和 2.00，远优于当前最先进的方法 GraphIQA [23] （SRCC 平均排名为 3.50，PLCC 平均排名为 3.75）和 TReS [5] （SRCC 平均排名为 4.00，PLCC 平均排名为 3.75）。所有这些结果都表明，与对比方法相比，我们所提出的 PMT-IQA 模型取得了最佳的整体性能。

为验证所提出的 PMT-IQA 模型的泛化能力，我们进一步开展了多项跨数据库测试。在这些测试中，使用一个特定数据集作为训练集，并使用另一个不同的数据集作为测试集。我们选取了四种具有竞争力的方法作为基线方法进行性能比较，包括 PQR、DB-CNN、HyperIQA 和 TReS。跨数据库测试包括四种设置：（1）在 LIVE-C 上训练，在 BID 上测试；（2）在 BID 上训练，在 LIVE-C 上测试；（3）在 LIVE 上训练，在 CSIQ 上测试；（4）在 CSIQ 上训练，在 LIVE 上测试。前两种设置针对真实失真，后两种针对合成失真。表 3 展示了对比方法和我们所提出的 PMT-IQA 模型的 SRCC 值。从结果可以看出，在全部四项跨数据库测试中，所提出的 PMT-IQA 模型均显著优于对比方法。

表 2. 各种方法在 BID、LIVE-C、LIVE 和 CSIQ 数据集上的 SRCC 和 PLCC 值，以及各方法的 SRCC 和 PLCC 平均排名。最佳结果以粗体标示，括号内的数字表示相应排名。我们报告十次运行中 SRCC 和 PLCC 的中位数。

	BID		LIVE-C		LIVE		CSIQ		平均排名	
方法	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
BRISQUE [18]	0.562(11)	0.593(11)	0.608(9)	0.629(13)	0.939(12)	0.935(11)	0.746(14)	0.829(10)	11.50(13)	11.25(12)
AlexNet [11]	-	-	0.766(10)	0.807(11)	0.932(13)	0.841(15)	0.766(13)	0.811(13)	12.00(14)	13.00(14)
ResNet50 [6]	0.583(10)	0.599(10)	0.824(9)	0.868(7)	0.947(10)	0.913(12)	0.823(9)	0.876(9)	9.50(9)	9.50(9)
ILNIQE [29]	0.516(13)	0.554(13)	0.432(11)	0.508(15)	0.903(14)	0.865(14)	0.806(11)	0.808(14)	12.25(15)	14.00(15)
HOSA [26]	0.721(9)	0.736(9)	0.640(8)	0.678(12)	0.946(11)	0.947(10)	0.741(15)	0.823(11)	10.75(11)	10.50(10)
BIECON [8]	0.539(12)	0.576(12)	0.595(10)	0.613(14)	0.961(8)	0.962(7)	0.815(10)	0.803(15)	10.00(10)	12.00(13)
SFA [15]	0.826(7)	0.840(7)	0.812(10)	0.833(10)	0.883(15)	0.895(13)	0.796(12)	0.818(12)	11.00(12)	10.50(10)
PQR [27]	0.775(8)	0.794(8)	0.857(3)	0.872(5)	0.965(6)	0.951(9)	0.873(8)	0.901(8)	6.25(7)	7.50(8)
DB-CNN [30]	0.845(6)	0.859(6)	0.851(5)	0.869(6)	0.968(5)	0.971(2)	**0.946(3)**	**0.959(1)**	4.75(6)	3.75(2)
HyperIQA [22]	0.869(3)	0.878(3)	0.859(2)	0.882(3)	0.962(7)	0.966(6)	0.923(5)	0.942(5)	4.25(4)	4.25(5)
UNIQUE [31]	0.858(5)	0.873(4)	0.854(4)	0.890(2)	0.969(2)	0.968(4)	0.902(7)	0.927(7)	4.50(5)	4.25(5)
CONTRIQUE [17]	-	-	0.845(7)	0.857(9)	0.960(9)	0.961(8)	0.942(5)	0.955(3)	7.00(8)	6.67(7)
GraphIQA [23]	0.860(4)	0.870(5)	0.845(7)	0.862(8)	**0.979(1)**	**0.980(1)**	0.947(2)	**0.959(1)**	3.50(2)	3.75(2)
TReS [5]	0.872(2)	0.879(2)	0.846(6)	0.877(4)	0.969(2)	0.968(4)	0.922(6)	0.942(5)	4.00(3)	3.75(2)
PMT-IQA（所提出的方法）	**0.874(1)**	**0.894(1)**	**0.866(1)**	**0.893(1)**	0.969(2)	0.971(2)	**0.949(1)**	0.951(4)	**1.25**(1)	**2.00**(1)

表 3.SRC C 在跨数据库测试中的比较。最佳结果 结果以粗体显示。

训练集	测试集	PQR [27]	DB-CNN [30]	Hyper-IQA [22]	TReS [5]	PMT-IQA（所提出的方法）
LIVE-C	BID	0.714	0.762	0.756	0.870	**0.897**
BID	LIVE-C	0.680	0.725	0.770	0.765	**0.782**
LIVE	CSIQ	0.719	0.758	0.744	0.738	**0.766**
CSIQ	LIVE	0.922	0.877	0.926	0.932	**0.934**

4.3 消融研究

我们在四个测试数据集上开展了若干细致的消融研究，以进一步验证多尺度特征提取模块和 PMT 模块的有效性。各变体如下：

(1) **ResNet**：基于 ImageNet 预训练的 ResNet50 架构，并添加全连接层进行预测（即不使用 MS 和 PMT）。(2) **Type1**：图 2 中的完整架构，仅使用 MS（即不使用 PMT）。(3) **Type2**：图 2 中的完整架构，使用 MS 和 PMT，并采用固定的 λ_1和 λ_2；我们根据测试实验在实现中使用 $\lambda_1 = \lambda_2 = 0.5$ 。(4) **PMT-IQA**：图 2 中的完整 PMT-IQA 架构，使用 MS 和 PMT，并采用式 10 中的动态任务权重。

为确保各变体的参数数量与完整的 PMT-IQA 相近，我们对隐藏维度进行了调整，以排除模型复杂度带来的性能提升，确保比较公平。如图4所示，与其他三个变体相比，PMT-IQA 在全部四个测试数据集上均取得了最佳性能。此外，MS 和 PMT 模块都能提升模型性能。与 MS 相比，在 LIVE、LIVE-C 和 CSIQ 上进行 PLCC 评估时，PMT 对整体性能的提升更为显著。此外，Type2 与 PMT-IQA 的结果对比表明

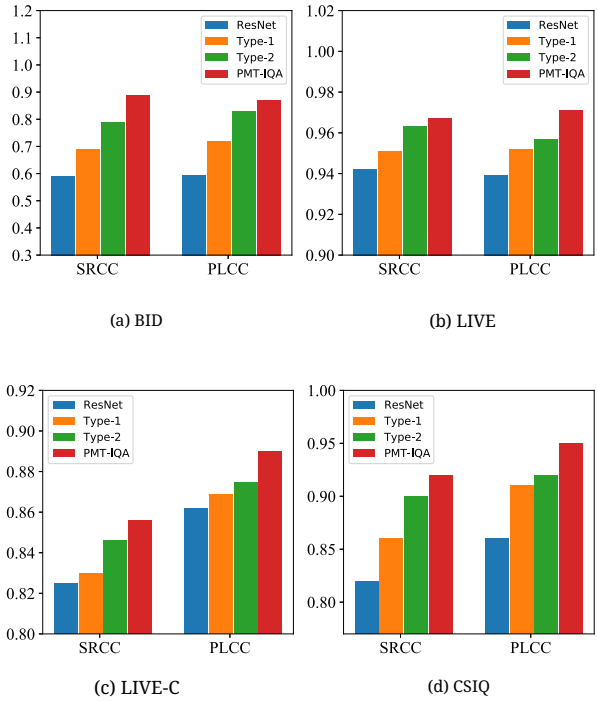


图4. 在 BID、LIVE、LIVE-C 和 CSIQ 数据集上的消融研究。

动态调整任务权重、使网络从简单任务逐步学习到复杂任务的策略是有效的。PMT-IQA 中新颖的任务渐进式切换对于训练阶段至关重要。因此，消融研究结果再次验证了所提出的 PMT-IQA 的有效性。

5 结论

本文提出了一种简单而有效的盲图像质量评价渐进式多任务学习模型。该模型包含一个多尺度特征提取模块和一个渐进式多任务学习模块，通过模拟由易到难的人类学习规律，帮助模型学习复杂的失真模式。大量实验结果表明，尽管所提出的 PMT-IQA 方法架构相对简单，但与各种基线方法相比，它在所有数据集上仍能取得更优或具有竞争力的表现。

致谢。本工作得到了国家自然科学基金委员会项目（编号：61866031）的资助。

参考文献

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna：一种新一代超参数优化框架。载于：第25届ACM SIGKDD知识发现与数据挖掘国际会议论文集。第2623–2631页（2019） 2. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: 课程学习。载于：第26届机器学习年度国际会议论文集。第41–48页（2009） 3. Ciancio, A., da Silva, E.A., Said, A., Samadani, R., Obrador, P., 等：基于多特征分类器的数字图像无参考图像模糊度评估。IEEE图像处理汇刊 **20**(1)，64–75（2010） 4. Ghadiyaram, D., Bovik, A.C.: 主观与客观图像质量的大规模在线众包研究。IEEE图像处理汇刊 **25**(1)，372–387（2015） 5. Golestaneh, S.A., Dadsetan, S., Kitani, K.M.: 通过Transformer、相对排序和自洽性进行无参考图像质量评估。载于：IEEE/CVF冬季计算机视觉应用会议论文集。第1220–1230页（2022） 6. He, K., Zhang, X., Ren, S., Sun, J.: 用于图像识别的深度残差学习。载于：IEEE计算机视觉与模式识别会议论文集。第770–778页（2016） 7. Kang, L., Ye, P., Li, Y., Doermann, D.: 通过多任务卷积神经网络同时估计图像质量与失真。载于：2015年IEEE国际图像处理会议（ICIP）。第2791–2795页。IEEE（2015） 8. Kim, J., Lee, S.: 全深度盲图像质量预测器。IEEE信号处理精选主题期刊 **11**(1)，206–220（2016） 9. Kim, J., Zeng, H., Ghadiyaram, D., Lee, S., Zhang, L., Bovik, A.C.: 用于图像质量预测的深度卷积神经网络模型：数据驱动图像质量评估的挑战与解决方案。IEEE信号处理杂志 **34**(6)，130–141（2017） 10. Kingma, D.P., Ba, J.: Adam：一种随机优化方法。arXiv预印本 arXiv:1412.6980（2014） 11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: 使用深度卷积神经网络进行ImageNet分类。神经信息处理系统进展 **25**（2012） 12. Krogh, A., Hertz, J.: 简单的权重衰减可以改善泛化。神经信息处理系统进展 **4** （1991） 13. Larson, E.C., Chandler, D.M.: 最显著失真：全参考图像质量评估与策略的作用。电子成像期刊 **19**(1)，011006（2010） 14. Lehman, A.: JMP基础单变量与多变量统计：分步指南。SAS研究所（2005） 15. Li, D., Jiang, T., Lin, W., Jiang, M.: 哪一种具有更好的视觉质量：晴朗的蓝天还是模糊的动物？IEEE多媒体汇刊 **21**(5)，1221–1234（2018）

16. Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., Zuo, W.: 使用深度神经网络进行端到端盲图像质量评估。IEEE图像处理汇刊 **27**(3)，1202–1213（2018） 17. Madhusudana, P.C., Birkbeck, N., Wang, Y., Adsumilli, B., Bovik, A.C.: 使用对比学习进行图像质量评估。IEEE图像处理汇刊 **31**，4149–4161（2022） 18. Mittal, A., Moorthy, A.K., Bovik, A.C.: 空间域中的无参考图像质量评估。IEEE图像处理汇刊 **21**(12)，4695–4708（2012） 19. Pavlov, I.: 条件反射。Les Etudes Philosophiques **17**(4)（1927年） 20. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: 对近期全参考图像质量评估算法的统计评估。IEEE图像处理汇刊**15**(11)，3440–3451（2006） 21. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout：一种防止神经网络过拟合的简单方法。机器学习研究期刊 **15**(1)，1929–1958（2014） 22. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: 在自适应超网络的引导下对真实场景中的图像质量进行盲评估。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第3667–3676页（2020） 23. Sun, S., Yu, T., Xu, J., Zhou, W., Chen, Z.: GraphIQA：学习用于盲图像质量评估的失真图表示。IEEE多媒体汇刊（2022） 24. Wang, X., Chen, Y., Zhu, W.: 课程学习综述。IEEE模式分析与机器智能汇刊 **44**(9)，4555–4576（2021） 25. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: 图像质量评估：从误差可见性到结构相似性。IEEE图像处理汇刊**13**(4)，600–612（2004） 26. Xu, J., Ye, P., Li, Q., Du, H., Liu, Y., Doermann, D.: 基于高阶统计聚合的盲图像质量评估。IEEE图像处理汇刊 **25**(9)，4444–4457（2016） 27. Zeng, H., Zhang, L., Bovik, A.C.: 采用概率质量表示的盲图像质量评估。载于：2018年第25届IEEE国际图像处理会议（ICIP）。第609–613页。IEEE（2018） 28. Zhai, G., Min, X.: 感知图像质量评估：综述。中国科学：信息科学 **63**(11)，1–52（2020） 29. Zhang, L., Zhang, L., Bovik, A.C.: 特征增强的完全盲图像质量评估器。IEEE图像处理汇刊 **24**(8)，2579–2591（2015） 30. Zhang, W., Ma, K., Yan, J., Deng, D., Wang, Z.: 使用深度双线性卷积神经网络进行盲图像质量评估。IEEE视频技术电路与系统汇刊 **30**(1)，36–47（2018） 31. Zhang, W., Ma, K., Zhai, G., Yang, X.: 在实验室和真实场景中进行不确定性感知盲图像质量评估。IEEE图像处理汇刊**30**，3474–3486（2021）