基于先验知识引导的彩色图像隐写分析架构搜索方法

李秋实[1]，谭舜泉[1]，李斌[2]，黄继武[1]
（1.深圳北理莫斯科大学工程系广东省智能感知与计算重点实验室，深圳 518116；2.深圳大学电子与信息工程学院，深圳 518060）

摘要：针对彩色图像隐写分析网络设计严重依赖人工经验、耗时且模型日益复杂的问题，提出了一种两阶段可微隐写分析架构搜索框架，以自动设计高效的网络架构。鉴于预处理层的关键作用，第一阶段构建了面向颜色通道的专用搜索空间，利用分组卷积及先验知识引导优化噪声残差提取，并证实其相比现有方法具有更优的高通特性和搜索效率。第二阶段在此基础上，进一步对网络的操作类型与深度进行联合搜索。在ALASKA-v2数据集上的大量实验表明，该方法搜索到的架构具有更小的模型尺寸，且在空域和JPEG域的检测性能均显著优于现有的手工设计及自动搜索的最先进模型。

关键词：信息隐藏；图像隐写分析；隐写；神经架构搜索；卷积神经网络
中图分类号：TP309
文献标志码：A
doi：10.11959/j.issn.1000

Prior knowledge guided architecture search for color image steganalysis

Li Qiushi[1], Tan Shunquan[1], Li Bin[2], Huang Jiwu[1]
1. Guangdong Key Laboratory of Machine Perception and Intelligent Computing, Faculty of Engineering, Shenzhen MSU-BIT University, Shenzhen 518116, China
2. College of Electronic and Information Engineering, Shenzhen University, Shenzhen 518060, China

Abstract: To address the issues of heavy reliance on manual expertise, high time consumption, and increasing model complexity in color image steganalysis network design, a two-stage differentiable steganalysis architecture search framework was proposed to automatically design efficient network architectures. In view of the critical role of preprocessing layers, the first stage constructs a specialized search space oriented towards color channels. It utilizes group convolution and prior knowledge to guide the optimization of noise residual extraction, demonstrating superior high-pass characteristics and search efficiency compared to existing methods. Building upon this foundation, the second stage performs a joint search for the network's operation types and depth. Experimental results demonstrate that the architectures derived by STARS outperform state-of-the-art models on the ALASKA-v2 dataset, achieving superior detection performance in both spatial and JPEG domains with significantly reduced model size.

Key words: information hiding, image steganalysis, steganography, neural architecture search, convolutional neural network

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：谭舜泉，tansq@smbu.edu.cn
基金项目：国家自然科学基金资助项目（No. 62272314，No. U23B2022，No. U22B2047）；深圳市基础研究项目（No. JCYJ20250604181211016，No.SYSPG20241211174032004）
Foundation Items：The National Natural Science Foundation of China (No. No. 62272314，No. U23B2022，No. U22B2047), Shenzhen Fundamental Research Project (No.JCYJ20250604181211016，No.SYSPG20241211174032004)

0 引言

隐写术是一种将秘密信息嵌入到数字媒体中而不引起感官怀疑的隐蔽通信技术，其潜在的安全风险日益凸显。作为隐写术的对立面，隐写分析旨在检测数字媒体中是否存在隐蔽信息，进而阻断非法的隐蔽通信。在当前的社交网络和即时通讯应用中，海量的彩色图像数据流使得隐蔽通信的风险急剧增加。实际的监管场景往往要求检测系统具备高吞吐量处理能力，并能够适应在资源受限的边缘计算节点或移动终端上进行实时筛查。

近年来，深度学习在隐写分析中的应用打破了以“富模型”[1]为代表的传统方法的发展瓶颈，显著提升了检测性能。然而，现阶段的深度隐写分析研究大多聚焦于灰度图像[2]。尽管彩色图像在实际应用中更为广泛，但由于其高维特性及颜色通道间复杂的相关性，跨通道的信号干扰使得极微弱隐写特征的提取难度增加，导致针对彩色图像的深度隐写分析研究仍相对滞后。现有的彩色图像隐写分析模型大多依赖专家经验进行手工设计，为避免微弱高频信号的丢失，通常保持特征图的高分辨率并堆叠大量滤波器[3-4]，或是直接迁移经过大规模预训练的通用视觉模型[5-8]。这些范式虽然取得了较好的检测精度，但普遍存在结构复杂、参数冗余巨大或过度依赖昂贵计算资源的问题，严重限制了其在大规模实时检测和边缘场景中的适用性。尽管已有部分研究尝试利用模型压缩技术对现有网络进行轻量化处理[9-11]，但这些后处理方法本质上受限于原始手工架构的固有拓扑结构，不可避免地会导致检测性能的显著损失。因此，探索一种能够突破人工试错局限、提高模型设计效率，并在保证检测精度的同时显著降低模型复杂度、满足实际部署边界条件的轻量级架构，显得尤为迫切。

为了顺应这一迫切需求，神经架构搜索（neural architecture search, NAS）技术提供了一种理想的解决路径[12-15]。作为一种数据驱动的自动化设计范式，NAS能够有效探索庞大的架构空间。然而，现有的NAS方法在图像隐写分析任务中仍面临极大的挑战。这主要归因于隐写分析与传统计算机视觉任务之间的“领域鸿沟”：隐写信号极其微弱且隐藏在图像的高频区域中，要求网络在早期阶段必须保持特征图的高分辨率。这种任务特性导致直接应用通用NAS方法时，会面临搜索空间巨大、显存占用极高以及搜索效率低下的算力瓶颈。而现有为数不多的隐写分析NAS探索，又往往因为缺乏有效的先验知识引导或搜索空间构建不合理，难以在计算代价和检测性能之间取得最佳平衡。

针对上述挑战，本文提出了一种基于先验知识引导的彩色图像隐写分析两阶段可微架构搜索框架，称为STARS（STeganalysis ARchitecture Search）。该框架旨在自动化地设计出既轻量又高效的隐写分析网络。本文的主要贡献总结如下。

（1）结合彩色图像隐写分析的特点，设计了一个紧凑的骨干网络，并构建了分层搜索空间。该空间涵盖了针对颜色通道的预处理操作以及后续的噪声提取和特征降维操作，特别关注颜色通道间的相关性与噪声残差的有效提取。

（2）提出了两阶段渐进式搜索策略。考虑到预处理层对隐写分析性能的关键影响，第一阶段专注于搜索最优预处理层；第二阶段在基于第一阶段导出的预处理层的基础上搜索后续网络结构。这种分步策略有效降低了搜索空间的复杂度，提高了搜索效率。

（3）在第一阶段搜索中引入了先验知识引导机制。利用SRM滤波器作为“教师”指导预处理层的搜索，通过知识蒸馏的方式帮助网络快速学习噪声残差提取能力。与直接使用高通滤波器初始化预处理层的策略相比，这种先验知识引导策略不仅加速了搜索收敛，还显著提高了最终搜索架构的检测性能和稳定性。

1 相关工作

本文探索面向深度学习彩色图像隐写分析的NAS方法。本节先介绍彩色图像深度隐写分析的架构演进，然后介绍当前NAS在隐写分析任务中的应用。

随着彩色图像在互联网上的广泛应用，针对彩色图像的隐写分析面临着更高的要求。研究者逐渐意识到直接将灰度网络[16-21]应用于彩色图像往往会忽视颜色通道间的相关性。为此，Zeng 等[3]提出了首个针对彩色图像的深度隐写分析网络WISER-Net，该网络引入了SRM滤波器组初始化的通道卷积（Channel-wise Convolution），通过“先分别处理颜色通道，后聚合特征”的策略，有效规避了普通卷积由于求和计算导致的“线性合谋攻击”。受

此设计启发，SRNet[16]等原用于深度学习灰度图像隐写分析的经典架构通过“颜色通道先拆分处理，后特征拼接”的改进方案，实现了对彩色图像的初步适配。针对空域与 JPEG 域的统一检测需求，UCNet[4]进一步发展了颜色通道预处理范式，通过引入固化的多域滤波器组（SRM 和 Gabor）来强化网络对高频噪声残差的捕捉能力。除了上述直接构建专用模型的方法，另一类研究思路则是借助在计算机视觉任务中预训练的基础模型（如 EfficientNet），利用其强大的表征能力进行迁移适配[5-8]。这类方法的内在机理在于通过“去语义化”改造，如去掉基础架构中底层卷积层的下采样操作以维持特征图的全分辨率，从而强化网络捕捉微弱高频残差特征的能力。尽管上述网络在检测精度上取得了长足进步，但它们仍未能摆脱人工启发式设计的固有局限。由于微弱的隐写信号往往难以捕捉，人工设计的隐写分析框架容易陷入“防御性扩张”的误区，即通过盲目保持高分辨率和堆叠大量滤波器来增强信号提取能力，导致网络架构往更宽、更深的方向发展，并伴随着严重的计算冗余。为了缓解这一痛点，近年来研究者开始探索模型压缩技术在隐写分析中的应用。Tan 等人率先将网络剪枝技术适配于隐写检测架构，通过移除冗余通道来提高模型效率[9-10]；随后，该团队进一步提出了分层张量分解算法，配合标准化的失真度量方法，实现了对隐写分析网络的无监督自动化压缩[11]。然而，这些技术在本质上是对已定型隐写分析架构的“后处理修剪”，其优化上限被原始手工设计的拓扑结构所限制，无法从源头上发现具备原生高效特征表达能力的轻量级拓扑。

为了突破人工设计的局限，研究人员开始尝试将自动化设计神经网络的 NAS 技术[12-15] 引入隐写分析领域，但面临着效率与性能难以兼顾的困境。例如，Yang 等人[22]提出了基于强化学习的 JS-NAG 方法，其需要耗费长达 45 个 GPU 天才能得到最终架构，极端的高昂计算成本限制了其可用性。随后，Deng 等人[23]提出了一种基于 PC-DARTS[15]的空域隐写分析架构搜索方法，虽然将搜索成本降低至 60 个 GPU 小时，但由于缺乏合适的先验引导，其搜索所得架构的检测性能仍低于手工设计的 SRNet。这主要归因于当前隐写分析 NAS 往往直接沿用计算机视觉领域的搜索空间，而这种空间存在显著的局限性：1）隐写分析旨在捕捉图像中极其微弱的高频残差，这要求网络在底层卷积层避免下采样以保持高分辨率，致使搜索过程中的超网特征图尺寸远超普通计算机视觉任务，引发严重的显存与算力消耗；2）计算机视觉任务中的搜索空间多采用预定义模块的重复堆叠，这种刻板的结构范式严重限制了特征提取的多样性，难以满足隐写分析对精细化残差建模的需求；3）由于缺乏足够的领域知识引导，容易导致生成的架构较臃肿，在推理延迟和计算量（FLOPs）方面不够友好。

鉴于此，本文提出了一种融合隐写分析领域先验知识的自动搜索框架（STARS）。该框架致力于重构契合微弱隐写分析特征提取逻辑的搜索空间，并探索化解全分辨率算力瓶颈的优化路径，以此填补现有方法中难以兼顾高检测精度与轻量化的研究空白。

2　本文方法

2.1　问题定义

本文将神经架构搜索问题形式化为一个双层优化问题。假设搜索空间 $\mathcal{A}$ 包含大量潜在的隐写分析架构，目标是从中找到最优架构 $a^* \in \mathcal{A}$，使得该架构在训练好权重 w_a^* 后，在隐写分析任务上性能最优。具体数学表达如下：

$$\begin{aligned} a^* &= \min_a L_{val}\left(w_a^*, a\right), \\ s.t.\quad w_a^* &= \arg\min_{w_a} L_{train}\left(w_a, a\right), a \in \mathrm{A\#}, \end{aligned} \tag{1}$$

其中，L_{train} 和 L_{val} 分别表示训练损失和验证损失，w_a 表示架构 a 对应的网络权重。该优化问题的核心在于最小化验证损失函数 $L_{val}\left(w_a^*, a\right)$ 来寻找和优化最优子网架构 a^*。对应的子网络权重 w_a^* 通过最小化训练损失 $L_{train}\left(w_a, a\right)$ 获得。在本文中，架构 a 由操作级参数 α 和深度级参数 β 共同表示。

2.2　搜索空间

为了在有限的计算资源下高效搜索出高性能的隐写分析网络，本文首先设计了一个紧凑且适配隐写分析任务的骨干网络，并基于此构建出分层搜索空间。本文所提出的 STARS 总体框架如图 1 所示。

不同于计算机视觉任务中常用的骨干网络，隐写分析网络需要尽早提取并保持高分辨率的噪声残差特征，避免过早下采样导致微弱隐写信号的丢失[16]。本文提出的骨干网络包含四个主要部分。

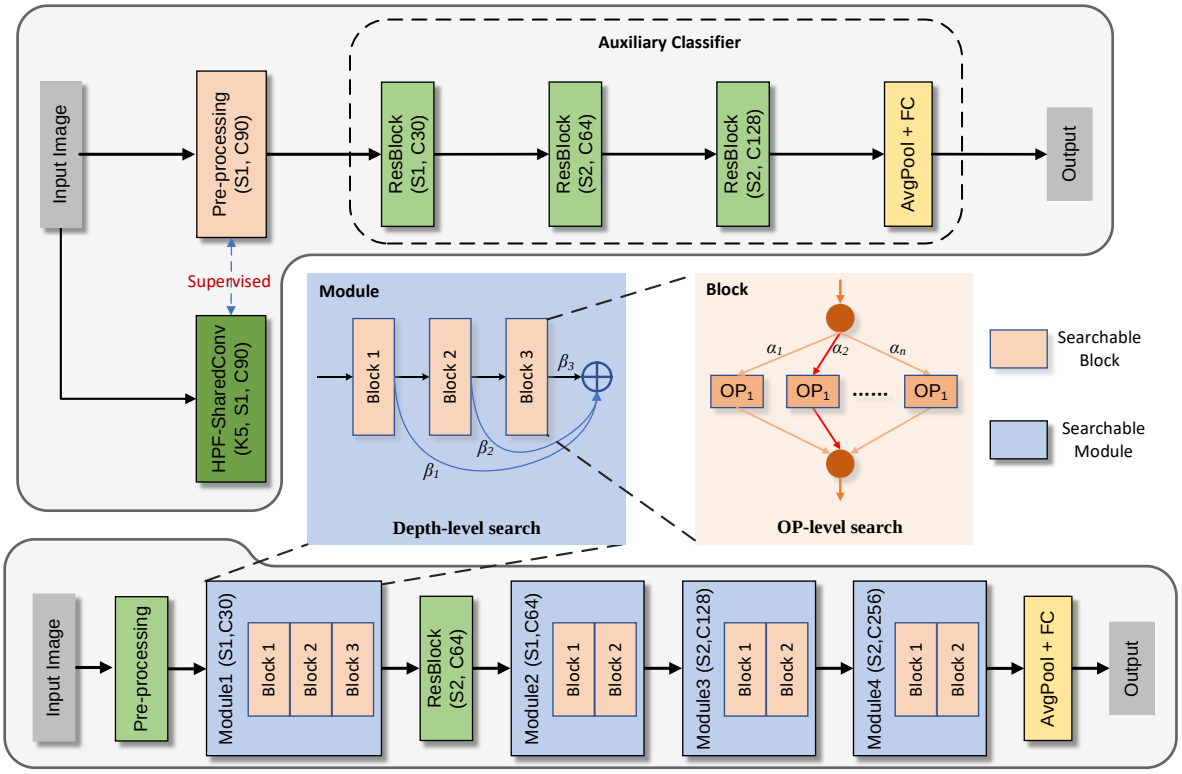


图 1　本文提出的 STARS 的总体框架

（1）预处理模块：作为网络的预处理层（也可被称为“茎干”层，Stem layer），负责对输入的彩色图像进行颜色通道预处理，提取初步的噪声残差。该层包含 1 层通道卷积（Channel-wise Conv），输出通道数设定为 90，且不涉及任何下采样操作，以保留丰富的残差信息。

（2）残差提取模块：由 3 层 3×3 残差卷积块（ResConv）组成，负责进一步提取和增强噪声特征，同时保持特征分辨率不变。为降低计算开销，该模块输出通道数被压缩至 30。

（3）特征压缩模块：通过级联的卷积层逐步进行下采样（分辨率减半）并增加通道数（分别增至 64、128），以提取更高级的语义特征并紧致化特征表示。可对应图 1 中的 $\mathcal{N}_2$ 中间 ResBlock 及后续的 Module2-4。图中模块标注 S2 表明对应 Block 1 的卷积步长（Stride）为 2。

（4）分类模块：利用全局平均池化将特征图转换为特征向量，并输入全连接层进行二分类（隐写图像/载体图像）。

基于上述骨干网络，本文构建了一个包含两个超网络 $\mathcal{N}_1$ 和 $\mathcal{N}_2$ 的分层搜索空间，分别对应两阶段搜索过程。STARS 整体框架如图 1 所示。其括号内的字母 K、S 和 C 分别表示卷积核大小、步长和输出通道数。**第一阶段搜索空间**（$\mathcal{N}_1$）：专注于预处理层的搜索。预处理层对于彩色图像隐写分析至关重要，其作用类似于传统方法中的滤波器组。为了充分探索通道间的相关性处理方式，$\mathcal{N}_1$ 中的预处理层包含三种类型的候选操作：（1）普通卷积（NormalConv）：标准的卷积操作，同时处理空间和通道信息；（2）共享卷积（SharedConv）：各通道独立卷积，但所有输入通道共享同一组卷积核。这种操作参数量少，能有效提取通道内的共性噪声特征；（3）分组卷积（GroupConv）：通道独立卷积，每个输入通道使用独立的卷积核组。相比 SharedConv，它能更灵活地适应不同颜色通道的统计特性。每种操作均包含 3×3、5×5、7×7 三种核尺寸，共计 9 种候选操作。**第二阶段搜索空间**（$\mathcal{N}_2$）：它由一个导出自 $\mathcal{N}_1$ 的预处理层、四个可搜索模块和一个全连接层组成。包含一个 $\mathcal{N}_1$ 中导出的预处理层，4 个可搜索的模块。该阶段从操作类型层面和深度层面构建了一个分级搜索空间，对四个可搜索模块进行操作级和深度级的联合搜索：1）操作级搜索：每个可搜索块（Block）包含 5 种候选操作：Res_K3（3×3 残差块）[24]、MBInvRes_K3_E0/

E3、MBInvRes_K5_E0/E3（不同核尺寸K和扩展比E的MobileNetV2倒瓶颈残差块）[25]。这些操作旨在平衡特征提取能力与计算效率。为避免歧义，本文明确定义：E3表示扩展比为3的标准倒残差模块（即包含1 × 1升维卷积、深度卷积和1 × 1降维映射）；而E0则表示完全省去初始升维步骤的深度可分离卷积模块（即直接对输入进行深度卷积，随后通过1 × 1卷积进行通道映射），以提供一种极致轻量化的结构候选。2）深度级搜索：允许网络自适应地选择每个模块的深度（即包含的块数量）。对于一个包含N个潜在块的模块，搜索算法决定保留前D个块，跳过剩余的N - D块，从而实现网络深度的自动调节。在这一阶段，四个可搜索模块的最大深度分别设为3，2，2和2。这种两阶段解耦搜索策略旨在化解浅层全分辨率特征提取与深层模块联合优化时引发的显存溢出及梯度耦合难题。它不仅大幅降低了超网的搜索计算代价，还有效保障了全局架构的搜索效率。

2.3　基于先验知识引导的搜索

在第一阶段的搜索中，为了提高搜索的效率和稳定性，本文提出了一种基于先验知识引导的搜索策略。如图1中的超网$\mathcal{N}_1$所示，与“输入图像”相连的下支路表示所使用的先验知识模块HPF-SharedConv，被用于从每个颜色通道中进行噪声残差特征提取，从而实现抑制图像内容，增强隐写信号的信噪比的目的。该模块由共享卷积Shared-Conv、批归一化层和ReLU激活层构成。其中，共享卷积层中的卷积核由空域富模型SRM[1]初始化，包含了30个不同的高通滤波器。SRM高通滤波器组中包含的滤波器大小并不统一，既存在3×3大小的滤波器，同时也存在着5×5大小的滤波器。为了使这些滤波器能够统一集成到一个5×5大小的卷积核中，需要先将这些高通滤波器扩展至相同的大小。对于较小的滤波器，本文采取了边缘补零的方式统一将其扩展至5×5大小。

假设网络输入彩色图像$\mathbf{X}$，每个颜色通道表示为$\mathbf{X}_i$, $i \in \{1,2,3\}$，SRM滤波器组$\mathbf{H}$。本文使用SRM滤波器组来计算每个颜色通道的噪声残差特征：

$$R_i = X_i \otimes H, \quad i \in \{1,2,3\}, \#(2)$$

其中$\otimes$表示2维卷积操作。具体操作如图2所示。通过SRM滤波器组对每个输入颜色通道进行残差特征提取，从而得到三组残差特征。再将这些输出特征沿着通道维度连接起来，最后得到90个噪声残差特征图。值得注意的是，由于该先验知识作为第一阶段搜索中的教师分支，因此在指导预处理层搜索的过程中，该教师分支中的卷积核权重保持固定，不可被优化。

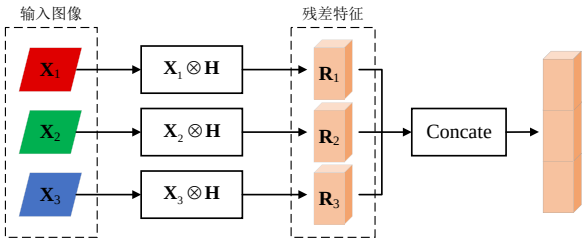


图2　彩色图像的先验特征提取操作

以往使用高通滤波器直接初始化预处理层的方法，如WISERNet[3]中使用的空域富模型SRM，以及UCNet[4]中使用的SRM+Gabor，可以看作是“硬指导”。在这种情况下，网络更容易陷入局部最优，且限制了网络的学习能力。相比之下，所使用SRM初始化的卷积层间接引导预处理层的搜索，帮助其快速学习噪声残差提取的相关知识，在保留更多隐写特征的同时过滤掉大量无关的图像内容，并且在超网$\mathcal{N}_2$中的预处理层将继承$\mathcal{N}_1$中预处理层学习到的权重。本文提出的方法，也就是使用了富模型SRM的先验知识的指导，可以看作是“软指导”。设$\mathcal{X} = \{x_i\}_{i=1}^n$是一组训练图像，$\mathcal{Y} = \{y_i\}_{i=1}^n$是相应的标签，$\hat{\mathcal{Y}} = \{\hat{y}_i\}_{i=1}^n$是$\mathcal{N}_1$对应的输出逻辑值。设$F^p(\cdot)$为$\mathcal{N}_1$中的预处理层，$F^k(\cdot)$为先验知识引导层。本算法引入$L_2$范数损失来缩小预处理层提取的特征图与先验知识引导层提取的噪声残差图之间的距离：

$$\frac{1}{n}\sum_{i=1}^{n} L_2\left(F^p(x_i), F^k(x_i)\right), \#(3)$$

其中L_2表示L_2范数损失。面向该隐写分析任务的损失函数为二元交叉熵损失：

$$L_{CE}(y_i, \hat{y}_i) = y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i). \#(4)$$

第一阶段的整体损失函数可表示为：

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\lambda \cdot L_2\left(F^p(x_i), F^k(x_i)\right) + L_{CE}(y_i, \hat{y}_i)\right\}, \#(5)$$

其中λ是超参数，其在实验中被设置为0.05。

2.4 搜索空间的连续松弛化

本节将介绍所提出的搜索策略，该策略依赖于对离散的搜索空间的连续化松弛，以此达到反向传播优化和架构选择的目的。

如第 2.1 节所述，给定整个搜索空间 $\mathcal{A}$，所提算法的目标是找到一个最优架构 $a^* \in \mathcal{A}$，使得在训练其权重 w_{a^*} 后，它可以获得最小的损失 $L_{val}\left(w_{a^*}, a^*\right)$。对于操作级搜索空间，设 $\mathcal{O}^l$ 为第 l 块中候选操作的集合，其中的候选操作已经在第 2.2 节中进行了详细的描述。第 l 块的输出公式可以被表示为：

$$x^{l+1} = \sum_{o \in \mathcal{O}^l} A_o^l \cdot o\left(x^l\right), \quad \text{s.t. } A_o^l \in \{0,1\}, o \in \mathcal{O}^l \qquad \#\#\#(6)$$

其中，A_o^l 表示二进制指示符，表示从第 l 块中的 $\mathcal{O}^l$ 中进行选择。在这种情况下，梯度反向传播不能用于解决二元变量的优化问题。为了以可微分的方式找到最优设计，DARTS[12]是首批将连续松弛化的方法引入 NAS 的研究之一，允许梯度反向传播通过 Softmax 在松弛的搜索空间中搜索架构。基于该策略，公式(6)中 A_o^l 可表达为:

$$A_o^l = \frac{\exp\left(\alpha_o^l\right)}{\sum_{o' \in \mathcal{O}^l} \exp\left(\alpha_{o'}^l\right)}, o \in \mathcal{O}^l, \#(7)$$

其中 α_o^l 是分配给第 l 块中的每个候选操作的连续架构参数。在进行了松弛化后，搜索空间 $\mathcal{A}$ 可以对操作级架构参数进行可微优化。

然而，当在搜索阶段中，块输出是候选操作的加权求和时，对计算资源的需求会随着候选操作的数量线性增加，正如 DARTS 所做的那样。为了解决这个问题，本工作采用了与[12]相同的思路。通过使用采样方法，只对块中某些候选操作的权重进行优化，这不仅节省了计算量，还提高了搜索效率。因此，本章引入了一种重参数化策略，称为 Gumbel-Softmax[26]：

$$x^{l+1} = \sum_{o \in \mathcal{O}^l} G_o^l \cdot o\left(x^l\right) = \sum_{o \in \mathcal{O}^l} \frac{\exp\left(\left(\alpha_o^l + g_o^l\right)/\tau\right)}{\sum_{o' \in \mathcal{O}^l} \exp\left(\left(\alpha_{o'}^l + g_{o'}^l\right)/\tau\right)} \cdot o\left(x^l\right), \quad \#(8)$$

$$\text{s.t. } \alpha_o^l = \log\left(\frac{\exp\left(\theta_o^l\right)}{\sum_{o' \in \mathcal{O}^l} \exp\left(\theta_{o'}^l\right)}\right), \#$$

其中 g_o 是从 Gumbel(0,1) 分布中独立同分布采样得到的随机变量，其生成方式为 $g = -\log(-\log u)$，其中 $u \sim \text{Uniform}(0,1)$。$\tau$ 是一个温度参数，用于控制新样本近似离散向量的接近程度。当 $\tau \to 0$ 时，Softmax 计算逐渐逼近 argmax，样本向量趋于独热向量（One-hot）。当 $\tau \to \infty$ 时，样本向量趋于均匀分布。θ_o^l 是与 α_o^l 对应的可学习参数，在本工作中，α_o^l 初始化为 0。值得注意的是，公式(8)中的嵌套结构旨在引入成熟的 Gumbel-Softmax 采样机制：内层操作负责将无约束参数 θ 归一化为合法的对数概率分布，外层则在此基础上进行可微离散采样。该重参数化机制避免了传统 DARTS 需同时计算所有候选操作导致的显存激增，在保证梯度连续可导的同时大幅降低了计算代价。

在操作级架构参数的优化中，以往的方法往往是简单地按 argmax 选取概率变量最大的操作进行训练。它可能更关注某些特定的操作，并且比其他操作更频繁地更新它们的权重，从而得到次优的架构。为了增加采样的随机性，使得在搜索的早期，更多的操作权重能够得到充分优化，所提算法分别采用了 Gumbel-Softmax 策略和均匀随机采样策略对两个子网进行采样，并同时通过梯度反向传播对采样子网络的权重进行优化。如图 1 中的“深度级搜索”模块所示，本算法为模块 s 中的第 l 个连接分配一个深度级架构参数 β_l^s。在搜索过程中，模块的输出是模块中每个块的输出特征图的加权求和。在导出架构时，每个模块中只选择一个候选连接。这意味着可以跳过相应块之后的所有块，从而有效地避免了架构冗余。通过 Softmax 连续松弛深度级架构参数，使其在搜索过程中也能以可微的方式进行优化：

$$x^{s+1}=\sum_{l\in\mathcal{S}}B_l^s\cdot x^s$$
$$=\sum_{l\in\mathcal{S}}\frac{\exp\left(\beta_l^s\right)}{\sum_{k\in\mathcal{S}}\exp\left(\beta_k^s\right)}\cdot x^s,\#\quad\#(9)$$

其中，x^s是第s个模块的输出，$\mathcal{S}^s$是第s个模块中的候选连接集。

2.5　搜索步骤

如第2.2节所述，为了增加NAS搜索空间的多样性，探索出更好的隐写分析架构，隐写分析架构的预处理层是必不可少的待搜索块之一。然而，根据隐写分析网络设计范例，在较低的层中不执行下采样，因此如果有非常多的候选操作，将会导致相当大的计算资源的需求。此外，预处理层与具体任务本身有着很强的相关性。因此，本文作者决定将预处理层与其他高级搜索空间分开搜索。所提出的STARS的伪代码如算法1所示。

算法1　所提出的两阶段STARS算法

输入　$\mathcal{D}_w$、$\mathcal{D}_a$

输出　最优架构a^*

1) 初始化网络权重w与架构参数α和β

// 第一阶段：预处理层搜索

2) while $\mathcal{N}_1$未收敛do

3) 从$\mathcal{D}_w$中采样批量数据d_t

4) 利用$\nabla_w L_{train}(w,\alpha)$在$d_t$上更新操作权重$w$

5) 从$\mathcal{D}_a$中采样批量数据d_v

6) 利用$\nabla_\alpha L_{val}(w,\alpha)$在$d_v$上更新架构参数$\alpha$

7) end while

8)根据$\mathcal{N}_1$中α分布，导出最优预处理层

// 第二阶段：操作级与深度级联合搜索

9) 将$\mathcal{N}_2$中的预处理层替换为在$\mathcal{N}_1$导出的预处理层10) while $\mathcal{N}_2$未收敛do

11) 从$\mathcal{D}_w$中采样批量数据d_t

12) 利用$\nabla_w L_{train}(w,\alpha,\beta)$在$d_t$上更新操作权重$w$

13) 从$\mathcal{D}_a$中采样批量数据d_v

14) 利用$\nabla_{\alpha,\beta} L_{val}(w,\alpha,\beta)$在$d_v$上更新架构参数$\alpha$和$\beta$

15) end while

16) 根据$\mathcal{N}_2$中α和β的分布，导出最有架构a^*

17) 在训练集上优化架构a^*

由于搜索空间以一种松弛化的方式得到连续表示，可以通过应用随机梯度下降来更新架构参数。该算法的搜索过程被分为两个阶段。首先，将训练集分成两个不相交的集合：$\mathcal{D}_w$和$\mathcal{D}_a$。在第一阶段，为了搜索到最优的预处理层，首先构造一个超网$\mathcal{N}_1$，由预处理层搜索空间和一个辅助分类器组成。然后，在富模型SRM的先验知识的指导下，交替地优化架构中的操作权重和采样架构参数。前者（操作权重）通过在$\mathcal{D}_w$上进行梯度下降$\nabla_w L_{train}(w,\alpha)$得到优化。后者（架构参数）通过在验证集$\mathcal{D}_a$上进行梯度下降$\nabla_\alpha L_{val}(w,\alpha)$从而得到优化。当$\mathcal{N}_1$收敛时，所提算法基于$\mathcal{N}_1$中的架构参数$\alpha$分布导出最佳预处理层，并舍弃辅助分类器。在第二阶段，通过继承第一阶段导出的预处理层的结构和经过优化的权重来构造超网$\mathcal{N}_2$，并继续搜索接下来的四个可搜索模块。采样子网的权重和$\mathcal{N}_2$中的架构参数（α和β）交替更新，直到$\mathcal{N}_2$收敛。最后，根据架构参数α和β在$\mathcal{N}_2$中的分布，导出最优架构。

3　实验结果和分析

在本节中，首先给出实验的实现细节。然后报告和分析本文提出的彩色图像隐写分析架构搜索算法的实验结果。最后，进行消融实验来评估该算法的有效性。

3.1　实验设置

1）**数据集与隐写算法**：为了评估所提出方法的有效性，实验选用了当前最权威的彩色图像隐写分析基准数据集：ALASKA-v2[5]。所使用图像分辨率为256×256，未压缩、以及压缩质量因子（Quality Factor，QF）为75和95（QF75，QF95）分别用于空域数据集和JPEG域数据集。随机选取20,000张图像，按14:1:5的比例划分为训练集、验证集和测试集。

为了构造隐写数据，在空域中，选取针对彩色图像设计的非加性隐写策略CMD-C[27]和GINA[28]作为攻击目标，并结合HILL[29]代价函数，嵌入载荷设为0.2和0.4 bpc（bits per channel pixel，每通道像素位数）。对于JPEG域，选取JUNIWARD[30]和JMiPOD[31]作为攻击目标，嵌入载荷设为0.2和0.4 bpnzac（bits per non-zero AC DCT coefficient，每非零交流DCT系数位数），并采用CCFR[32]作为跨通道载荷分配策略。

2）**评估指标**：本节实验从多个角度对所提出

的隐写分析模型进行评估，包括：准确率（ACC）、ROC 曲线下加权面积（wAUC）[5]，和虚警率为5%时的漏检率（MD5）[33]。所使用的模型复杂度指标，包括参数的数量（Params）、乘加运算数量（Multi-Adds）和在大小为256×256的单张彩色图像上的推理时间（Infer）。

3）**超参数设置**：在搜索过程中，训练集按7:3划分成两个子集，分别用于权重和架构参数的优化。对于权重w，使用Adamax优化器，初始学习率为0.001。对架构参数α和β采用Adam优化器，学习率为0.01。第一阶段进行120个轮次优化，第二阶段采用相同超参数，并将架构参数的优化推迟了20个轮次，以充分预热候选操作权重。所使用的批大小为32。公式(8)中的初始温度参数τ设为5.0，衰减因子为0.96。所有实验均在单张NVIDIA Tesla A100 GPU上完成。

为了比较隐写分析模型的检测结果，除非另有说明，对实验中所涉及到的对比模型，均在训练策略中引入了先进的无配对约束策略 （noPC）以保证实验的公平性和完整性。具体的策略将遵循最新公开可用的训练设置[4][8][16]，并使用相应模型所提供的源代码在相应的目标数据集上报告检测性能。在空域的相关实验中，当遇到收敛问题时，通过加载在更高有效载荷（0.4 bpc）的相应隐写术上训练好的模型参数，进行课程学习，以解决此问题。在JPEG域的相关实验中，图像在输入网络前均被解压到量化前YCbCr格式。所使用的数据增强策略仅包括翻转和旋转。为了解决模型在低嵌入率或高压缩质量等困难样本上难以收敛的问题，本实验采用了课程学习策略，即先在较易任务（高嵌入率0.4 bpc或低压缩质量QF75）上进行预训练，再将模型权重迁移至困难任务（0.2 bpc或QF95）上进行微调。具体而言，在空域实验中，仅针对0.2 bpc下出现收敛瓶颈的SRNet采用了该策略；而在JPEG域中，考虑到QF95的挑战性，所有对比模型均默认基于QF75的预训练模型进行微调。此外，本文报道的检测性能均为3个不同随机种子下的平均结果，多次搜索均稳定收敛于同一架构。在不同型号GPU（如RTX 4090）上的交叉验证进一步证实了所提算法具有高度的结构一致性与性能稳定性。

3.2 在ALASKA-v2-COLOR256数据集上的实验评估

本节将报告所提出的模型对上述隐写算法的检测性能，并将所提出的模型与目前先进的基于深度学习的隐写分析器在空域和JPEG域上进行比较。

1) 在空域上的性能：表1展示了SRNet[16]、UCNet[4]、SwT-SN[34]、TSNet[35]和所提算法搜索到的网络针对两种非加性隐写方案CMD-C-HILL和GINA-HILL的检测性能比较，有效载荷为0.4 bpc和0.2 bpc。具体而言，当在嵌入率为0.2 bpc的数据集上训练SRNet时遇到了收敛问题。为了解决这一问题，本实验在以0.4 bpc嵌入率训练的模型基础上采用了课程学习策略。结果表明，对于CMD-C-HILL和GINA-HILL，本实验搜索到的模型优于其他模型。尽管在0.2 bpc嵌入率下针对CMD-C-HILL时，本文提出的方法在MD5指标上略逊于UCNet约0.06%，但它仍展现出极具竞争力的性能。此外，由于SRNet、SwT-SN和TSNet最初是针对灰度图像设计的，本实验采用了UCNet所采用的策略将其适配于彩色图像。然而，即使进行了这些适配，与UCNet和STARS等专门为彩色图像设计的隐写分析模型相比，它们的性能仍有显著差距。

表1 本文提出的模型在ALASKA-v2-COLOR256空域上的检测性能比较。带下划线的结果表示引入了课程学习策略

隐写分析模型	CMD-C-HILL						GINA-HILL					
	0.4 bpc			0.2 bpc			0.4 bpc			0.2 bpc		
	ACC	wAUC	MD5	ACC	wAUC	MD5	ACC	wAUC	MD5	ACC	wAUC	MD5
SRNet	76.72	86.52	53.68	<u>65.60</u>	<u>77.35</u>	<u>78.92</u>	72.96	83.06	69.72	<u>66.84</u>	<u>77.84</u>	<u>86.72</u>
UCNet	84.47	95.87	27.38	78.30	91.92	**41.84**	82.73	94.83	32.08	75.24	88.93	48.84
SwT-SN	82.84	95.35	31.40	74.61	89.93	50.02	80.11	93.83	37.94	68.51	84.67	62.46
TSNet	77.49	91.77	44.92	63.75	78.96	71.28	82.08	94.49	33.96	64.90	80.39	71.66
STARS	**85.19**	**96.08**	**26.14**	**79.27**	**92.16**	41.90	**83.85**	**95.10**	**31.54**	**76.60**	**91.21**	**47.72**

2) 在 JPEG 域上的检测性能：对于在 JPEG 域上的检测性能对比研究，本节实验选择 SRNet 和 EfficientNet-B4-NS[8] 作为对比模型。具体来说，EfficientNet 模型已经在大规模数据集 ImageNet 上进行了预训练，而 EfficientNet-B4-NS 是对 EfficientNet-B4 网络，移除了其首层步长，也就是首层不进行下采样。这样的领域知识的引入通常会给 EfficientNet 带来相当大的性能提升。在 JPEG 域的 ALASKA-v2-COLOR256 数据集上的检测性能对比如表 2 所示。在 QF75 场景中，本文提出的方法展现了极具竞争力的性能，尤其在 ACC 上取得了显著优势，尽管 UCNet 在 0.4 bpnzac 和 0.2 bpnzac 嵌入率下的 wAUC 方面略优于本文所出题的模型。在公认具有挑战性的 QF95 场景中，所提出的 STARS 显著优于现有的最先进模型。总体而言，由于 UCNet 结合了来自空域和 JPEG 域的广泛先验知识（包括 SRM 和 Gabor 滤波器），使其能够在两个域中都表现出稳健且有效的性能。然而，这也导致了显著增加的计算负担。

3）模型大小和计算代价：表 3 展示了参数量（Params）, 乘加计算量（Multi-Adds）和推理时间（Infer）上的比较。在表中分别展示了所提出的算法在对抗 CMD-C-HILL（C）、GINA-HILL（G）和 QF75-JUNIWARD（JU）时分别搜索的模型。对于 Params，所提算法的三个搜索网络都在 0.3 M 左右，这比其他模型要小得多。EfficientNet-B4-NS 的参数量比所提算法搜索的网络多大约 50 倍。在推理时间方面，所提算法的搜索网络也比其他模型低得多。由于所提算法使用的深度学习框架提供了许多算法来加速矩阵和卷积计算以及硬件方面（如数据 I/O 等）的支持，因此网络的推理时间通常与计算量没有恒定的比例相关性。搜索到的架构如图 3 所示。为了表示不同的块操作，图中使用不同颜色和宽度的矩形。K*x*E*y* 表示卷积核大小 $x \times x$ 和扩展比 y，且每个块上均被标记了输出通道数，蓝线用于划分模块。

需要注意的是，在上述三种场景下得到的预处理层均为 GroupConv，而不是 UCNet 中采用的 SharedConv。尽管 GroupConv 相比 SharedConv 具有更多的参数，但它能够适应不同颜色通道的统计特性，以充分提取噪声残差。为了理解 SharedConv 和 GroupConv 提取噪声残差的能力，本节使用傅里叶变换将这两种操作的特征图转换到归一化的频域。在图 4 中，展示了 SharedConv 和 GroupConv 对傅里叶频率分量所表现出不同的行为。如图 4(c) 和图 4(d)所示，GroupConv 相比 SharedConv 减少了更多的低频分量。在图 4(e)中，傅里叶变换的特征图的相对对数幅值可以看出，虽然两种操作都可以放大高频信号，但 GroupConv 的效果要强于 SharedConv，这意味着从彩色图像隐写分析的角度来看，GroupConv 对信噪比的增强效果更好（图像内容是“噪声”，而隐写噪声是“信号”）。该结果进一步印证了第一阶段中融合先验知识指导的预处理层搜索策略的有效性。

4) 与其他 NAS 方法的对比：为了突出 STARS 具有更好的搜索能力，本节将其与最先进的隐写分析 NAS 方法 PC-DARTS-Deng[23] 进行了比较。PC-DARTS-Deng 也是一种可微分的 NAS 方法，但它只搜索少数复杂的单元模块，然后进行重复堆叠。如表 4 所示，当在 ALASKA-v2-COLOR256 上搜索时，所提出的 STARS 比 PC-DARTS-Deng 节省了 9.5 个 GPU 小时。表 5 显示了所提出的 NAS 方法与 PC-DARTS-Deng 在检测 CMD-C-HILL 和 CCFR-JUNIWARD 时的性能比较。如表所示，在空域和 JPEG 域中，所提算法的搜索架构比 PC-DARTS-Deng 导出的模型性能更好，消耗的存储和计算资

表 2　　本文提出的模型在 ALASKA-v2-COLOR256 的 JPEG 域上的检测性能比较

隐写分析模型	QF75, CCFR-JUNIWARD						QF95, CCFR-JUNIWARD					
	0.4 bpnzac			0.2 bpnzac			0.4 bpnzac			0.2 bpnzac		
	ACC	wAUC	MD5	ACC	wAUC	MD5	ACC	wAUC	MD5	ACC	wAUC	MD5
SRNet	89.20	94.56	32.88	73.60	85.46	65.40	73.32	86.29	63.72	59.62	71.06	86.16
UCNet	88.70	**97.09**	22.16	73.93	**87.40**	60.08	73.00	87.25	58.44	59.29	72.60	**82.96**
EfficientNet-B4-NS	85.11	96.62	22.92	72.05	86.53	**59.58**	70.81	85.43	62.60	58.31	70.95	85.40
STARS	**89.23**	96.63	**20.76**	**75.55**	87.37	61.76	**74.45**	**88.15**	**57.80**	**59.98**	**73.00**	83.16

表3　模型大小、计算代价和推理时间方面的比较

模型	Params (M)	Multi-Adds (B)	Infer (ms)
SRNet	4.778	6.033	4.84
UCNet	1.117	7.160	4.07
EfficientNet-B4-NS	17.552	7.268	4.07
STARS-C	**0.311**	**4.896**	**2.90**
STARS-G	0.342	5.377	3.38
STARS-JU	0.325	5.834	3.05

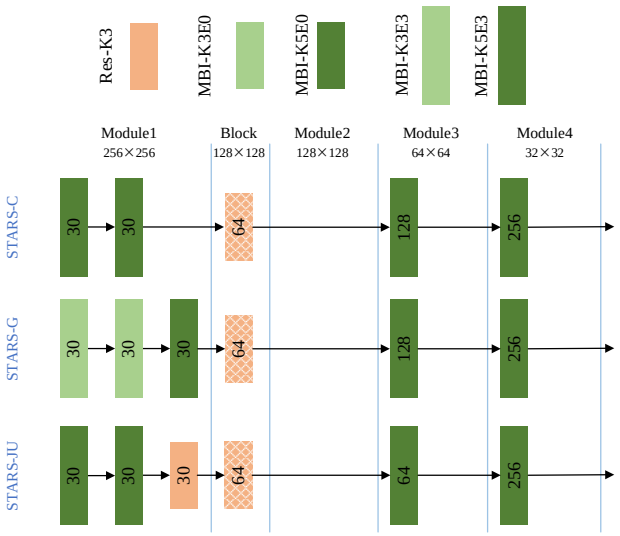
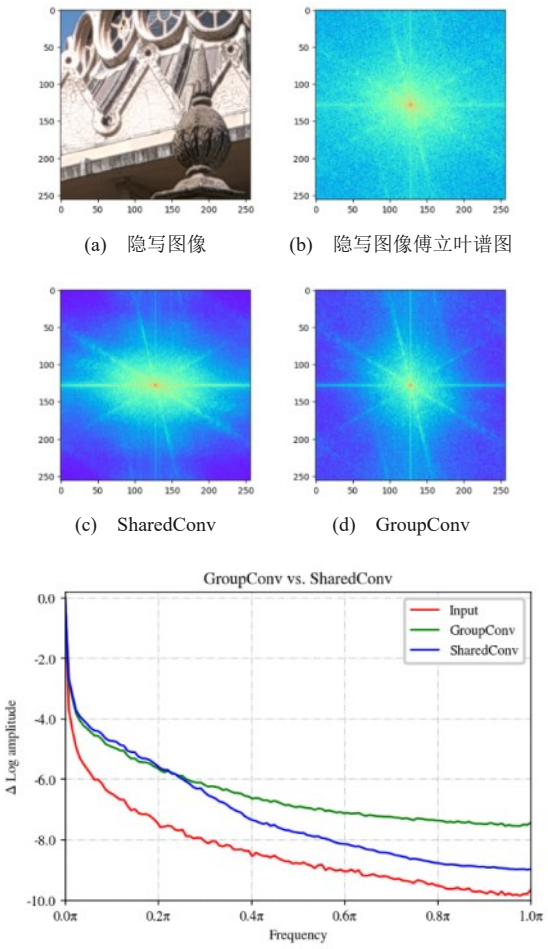


图3　最终导出的搜索架构的可视化



(a)　隐写图像　(b)　隐写图像傅立叶谱图

(c)　SharedConv　(d)　GroupConv

(e)　经过傅里叶变换的特征图的相对对数幅值

图4　对 SharedConv 和 GroupConv 的傅里叶分析

源更少。

3.3　模型泛化性评估

为了评估所提出的方法的泛化性，本节进行了跨隐写算法和跨图像分辨率的实验。首先，表6展示了跨隐写方法评估的结果，即使用CMD-C-HILL以0.4 bpc负载嵌入的ALASKA-v2-COLOR256数据集上训练的网络，在使用CMD-C-SUNIWARD[30]以0.4 bpc负载嵌入的同个数据集上进行评估。可以看出，与其他网络相比，所提出的模型表现出更好的性能。表7展示了跨图像分辨率的验证实验，

表4　在 ALASKA-v2-COLOR256 上与其他搜索方法的搜索成本比较

搜索算法	搜索空间	搜索代价（GPU 时）
PC-DARTS-Deng	基于单元格的搜索空间（cell-based）	45.5
STARS	基于块的搜索空间（block-wise）	**36**

表5　所提出的算法在 ALASKA-v2-COLOR256 和 PC-DARTS-Deng 的性能比较

目标隐写术	隐写分析架构搜索算法	Params (M)	Multi-Adds (B)	ACC	wAUC	MD5
CMD-C-HILL 0.4 bpc	PC-DARTS-Deng	0.724	17.908	83.62	93.10	31.14
	STARS	**0.311**	**4.896**	**85.19**	**96.08**	**26.14**
QF75, CCFR-JUNIWARD 0.4 bpnzac	PC-DARTS-Deng	0.538	5.711	87.32	95.76	24.52
	STARS	**0.325**	**5.834**	**89.23**	**96.62**	**20.76**

表 6　跨隐写方案的检测性能比较

隐写分析模型	CMD-C-HILL → CMD-C-SUNIWARD		
	ACC	wAUC	MD5
SRNet	69.38	82.06	73.36
UCNet	72.24	89.30	56.24
SwT-SN	73.29	90.15	53.52
TSNet	56.63	77.97	85.26
STARS	**74.03**	**91.36**	**52.46**

表 7　跨图像分辨率的检测性能的比较

隐写分析模型	256 → 512		
	ACC	wAUC	MD5
SRNet	76.72	86.56	46.92
UCNet	86.39	96.82	23.78
SwT-SN	83.55	95.94	29.10
TSNet	78.49	92.56	40.24
STARS	**86.63**	**97.09**	**22.02**

其中所有网络在使用 CMD-C-HILL 以 0.4 bpc 在 ALASKA-v2-COLOR256 上进行训练，并在相同的隐写方案下在 ALASKA-v2-COLOR512 上进行评估。可以看到，所搜索到的架构的性能优于其他网络，这意味着所搜索的架构表现出更高的泛化水平。需要指出的是，尽管 STARS 提升了跨隐写算法的泛化性（如表 6 所示），但在极低嵌入率或高强度压缩的极端场景下，完全从零开始搜索极微弱特征仍具挑战性，目前仍需依赖课程学习等策略辅助。这也是未来探索更加强健的先验引导机制的重要方向。

3.4　消融实验

为了进一步阐明深度级搜索和先验知识引导（PKG）策略的有效性，本节在使用 CMD-C-HILL 算法，以 0.4 bpc 负载进行嵌入的 ALASKA-v2-COLOR256 数据集上进行了一系列消融实验。实验结果见表 8。首先，为了验证深度级搜索的有效性，该消融实验在第二阶段不进行深度级搜索，在确定最终架构时保持初始块数量（STARS with fixed depth）。实验结果表明，与“STARS with fixed depth”相比，使用所提出的 STARS（表 8 中最后一行）搜索的架构大约是该模型（“STARS with fixed depth”）大小的三分之一，并且取得了相当的性能，这表明深度级搜索策略可以有效地减少架构的冗余，并且在增加了潜在架构多样性的同时更有可能找到性能更好的架构。在表 8 中，“one-stage”表示将所提出的二阶段 STARS 搜索算法改为一阶段 STARS 算法，从而使得预处理层和更高的模块在同一个超网中进行搜索。可以看出，所提出的二阶段 STARS 相比一阶段 STARS 均有更好的检测性能。在针对“PKG”的消融实验中，“w/o PKG”表示在搜索的第一阶段，不采用所提出的 PKG 策略，而是直接使用 SRM 滤波器初始化预处理层权重。而“w/ PKG”表示使用本文所提出的先验知识引导策略。从该消融实验的结果中，可以看到，本文所提出的 PKG 策略，能够有效提高隐写分析检测性能。不仅如此，不论是对于一阶段 STARS 还是如本文算法最终所采用二阶段 STARS，PKG 策略均有利于模型探索出性能更强的架构。

6　结束语

本文提出了一个二阶段可微分的神经架构搜索框架 STARS，用于在操作级和深度级两种维度中自动设计有效的彩色图像隐写分析的深度学习架构。首先提出了一种紧凑的彩色图像深度学习隐写分析骨干网络，并在此基础上进一步构建了分级搜索空间。在 SRM 的先验知识指导下，STARS 可以比其他隐写分析 NAS 框架更有效地搜索出轻量级的隐写分析架构。在 ALASKA-v2 数据集上针对空域和 JPEG 域的不同隐写方案对搜索的架构进行了评估，结果表明本文所提出的搜索模型比其他手工制作的最先进模型和自动搜索模型表现出更好的性能。未来工作中，将从以下两个方面提升隐写分析模型架构搜索算法的能力：1）引入更多样化的模块（例如注意力模块），以进一步开发性能更好的隐写分析架构；2）引入更多的搜索维度（例如每个块的通道数），以寻找更加轻量级的架构。

表 8 消融实验：验证深度级搜索和先验知识引导策略的有效性

搜索算法	Params (M)	Multi-Adds (B)	ACC	wAUC	MD5
STARS with fixed depth	0.899	6.817	84.79	92.68	29.72
One-stage STARS w/o PKG	0.301	4.271	78.19	91.84	51.18
One-stage STARS w/ PKG	0.852	5.475	80.32	92.77	43.34
Two-stage STARS w/o PKG	0.303	4.766	82.08	94.38	33.78
Two-stage STARS w/ PKG	0.311	4.896	**85.19**	**96.08**	**26.14**

参考文献：

[1] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images [J]. IEEE Transactions on Information Forensics and Security, 2012, 7 (3): 868-882.

[2] Luo W, Wei K, Li Q, et al. A comprehensive survey of digital image steganography and steganalysis[J]. APSIPA Transactions on Signal and Information Processing, 2024, 13(1): 1-67.

[3] Zeng J, Tan S, Liu G, et al. WISERNet: Wider separate-then-reunion network for steganalysis of color images[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(10): 2735-2748.

[4] Wei K, Luo W, Tan S, et al. Universal deep network for steganalysis of color image based on channel representation[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3022-3036.

[5] Cogranne R, Giboulot Q, Bas P. ALASKA#2: Challenging academic research on steganalysis with realistic images[C]//2020 IEEE International Workshop on Information Forensics and Security. 2020: 1-5.

[6] Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks[C]// Proceedings of Machine Learning Research: Vol. 97 Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019: 6105-6114.

[7] Yousfi Y, Butora J, Khvedchenya E, et al. ImageNet pre-trained CNNs for JPEG steganalysis[C]//2020 IEEE International Workshop on Information Forensics and Security. 2020: 1-6.

[8] Butora J, Yousfi Y, Fridrich J. How to pretrain for steganalysis[C]//Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security. Association for Computing Machinery, 2021: 143-148.

[9] Li Q, Shao Z, Tan S, et al. Non-structured pruning for deep-learning based steganalytic frameworks[C]//2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. 2019: 1735-1739.

[10] Tan S, Wu W, Shao Z, et al. CALPA-NET: Channel-pruning-assisted deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 131-146.

[11] Tan S, Li Q, Li L, et al. STD-NET: Search of image steganalytic deep-learning architecture via hierarchical tensor decomposition[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(3): 2657-2673.

[12] Liu H, Simonyan K, Yang Y. DARTS: Differentiable architecture search[C]//International Conference on Learning Representations. 2019.

[13] Dong X, Yang Y. Searching for a robust neural architecture in four GPU hours[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1761-1770.

[14] Hu Y, Wu X, He R. TF-NAS: Rethinking three search freedoms of latency-constrained differentiable neural architecture search[C]//Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 123-139.

[15] Xu Y, Xie L, Zhang X, et al. PC-DARTS: Partial channel connections for memory-efficient architecture search[C]//International Conference on Learning Representations. 2020.

[16] Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1181-1193.

[17] He J, Weng S, Yu L, Zhang C, Chen W. An image steganoganalyzer with comprehensive detection performance[J]. IEEE Signal Processing Letters, 2023, 30: 1682-1686.

[18] Weng S, Chen S, Yu L, Sun S. Lightweight and effective deep image steganalysis network[J] IEEE Signal Processing Letters, 2022, 29: 1888-1892.

[19] He J, Weng S, Yu L, Chen D. Steganalysis network with two-branch preprocessing for spatial and JPEG domains[J] IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(2): 1451-1463.

[20] Wang F, Fu Z, Zhang X, et al. Pair-wise confidence difference-based pseudo-label selection for universal mismatched steganalysis[C]//Proceedings of the 33rd ACM International Conference on Multimedia. 2025: 8673 – 8681.

[21] Li H, Luo X, Zhang Y, et al. A convolutional neural network steganalysis method based on ShuffleBottleneck and attention mechanism[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22 (6): 7389-7402.

[22] Yang J, Lu B, Xiao L, et al. Reinforcement learning aided network architecture generation for JPEG image steganalysis[C]//Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security. New York, NY, USA: Association for Computing Machinery, 2020: 23-32.

[23] Deng X, Luo W, Fang Y. Spatial steganalysis based on gradient-based neural architecture search [C]//Provable and Practical Security. Cham: Springer International Publishing, 2021: 365-375.

[24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[25] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.

[26] Maddison C J, Mnih A, Teh Y W. The concrete distribution: A continuous relaxation of discrete random variables[C]//International Confer-

ence on Learning Representations. 2017.

[27] Tang W, Li B, Luo W, et al. Clustering steganographic modification directions for color components[J]. IEEE Signal Processing Letters, 2016, 23(2): 197-201.

[28] Wang Y, Zhang W, Li W, et al. Non-additive cost functions for color image steganography based on inter-channel correlations and differences[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2081-2095.

[29] Li B, Wang M, Huang J, et al. A new cost function for spatial image steganography[C]//2014 IEEE International Conference on Image Processing. 2014: 4206-4210.

[30] Holub V, Fridrich J, Denemark T. Universal distortion function for steganography in an arbitrary domain[J]. EURASIP Journal on Information Security, 2014, 2014(1): 1-13.

[31] Cogranne R, Giboulot Q, Bas P. Efficient steganography in JPEG images by minimizing performance of optimal detector[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1328-1343.

[32] Taburet T, Filstroff L, Bas P, et al. An empirical study of steganography and steganalysis of color images in the JPEG domain[C]//Digital Forensics and Watermarking. Cham: Springer International Publishing, 2019: 290-303.

[33] Cogranne R, Giboulot Q, Bas P. The ALASKA steganalysis challenge: A first step towards steganalysis[C]//IH&MMSec'19: Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. New York, NY, USA: Association for Computing Machinery, 2019: 125-137.

[34] Weng S, Sun S, Yu L. Fast SwT-based deep steganalysis network for arbitrary-sized images[J]. IEEE Signal Processing Letters, 2023, 20: 1782-1786.

[35] Guo F, Sun S, Weng S, et al. A two-stream-network based steganalysis network: TSNet[J]. Expert Systems with Applications, 2024, 255: 124796.

李秋实（1996– ），男，博士，意大利佛罗伦萨大学博士后，主要研究方向为多媒体取证、隐写及隐写分析等。

谭舜泉（1980– ），男，博士，深圳北理莫斯科大学教授、博士生导师，主要研究方向为多媒体取证、隐写及隐写分析、深度学习等。

李斌（1982– ），男，博士，深圳大学教授、博士生导师，主要研究方向为多媒体信息安全、智能处理等。

黄继武（1962– ），男，博士，深圳北理莫斯科大学教授、博士生导师，主要研究方向为多媒体取证、信息隐藏、多媒体信号处理、模式识别等。