

Learning to Structure an Image with Few Colors

Yunzhong Hou, Liang Zheng, Stephen Gould

摘要

色彩和结构是构成图像的两大支柱。通常，通过丰富的颜色可以很好地表现结构，使神经网络能够识别图像中的对象。然而，在色彩空间的极端限制下，结构往往会消失，因此神经网络可能无法理解图像。作者对颜色和结构之间的这种相互关系感兴趣，故研究了如何在颜色空间的限制下识别和保留信息量最大的图像结构，使生成的图像可以高精度地被识别。为此，作者提出了一种名为 ColorCNN 的颜色量化网络，它以端到端的方式从分类任务损失中学习如何构建图像。给定颜色空间的大小，ColorCNN 通过生成颜色索引图和 RGB 调色板来量化原始图像中的颜色。然后，这个颜色量化的图像被输入到一个预训练的任务网络来评估它的性能。实验中，在仅使用 1 位颜色空间（即两种颜色）时，所提出的网络在 CIFAR10 数据集上实现了 82.1% 的 top-1 准确率，大大优于传统的颜色量化方法。对于应用程序，当使用 PNG 编码时，所提出的颜色量化方法在极低比特率范围内显示出优于其他图像压缩方法的优势。

关键词：颜色量化；图片压缩

1 引言

颜色和结构是图像的两个重要组成成分。结构被视为形状、纹理等的组合，与颜色密切相关。只有当存在足够的颜色时，结构才能很好地呈现。该文章感兴趣的是如何在颜色约束下最好地呈现结构。

与此密切相关的一系列研究工作是“颜色量化”。颜色量化研究如何在受限颜色空间中保持视觉相似性^[1-2]。这个问题以人为中心，因为它通常关注人类观看的视觉质量。特别是，大多数现有方法都是在相对较大的色彩空间（例如 8 位）下设计的，因此量化后的图像在视觉上仍然与原始图像相似。在较小的颜色空间中，例如 2 位或 1 位，颜色量化仍然是一个悬而未决的问题。

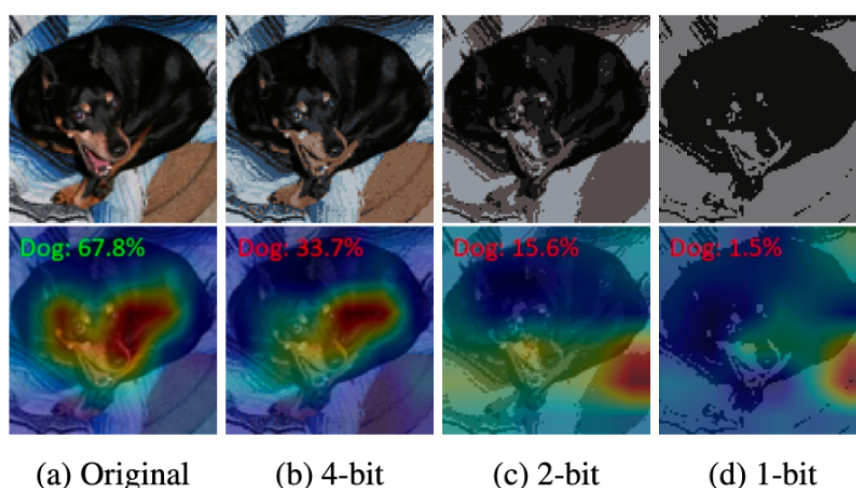


图 1: 上：传统方法颜色量化的结果；下：量化结果在预训练神经网络的分类器相应（class activation map, CAM）。绿色：正确识别；红色：识别失败。

自然图像通常包含丰富的色彩和结构。当色彩空间有限时，它们的连接会破坏结构。例如，在图 1 的第一行中，结构随着颜色空间的减少而消失。此外，作者认为在原始自然图像上训练的神经网络可能无法有效识别颜色较少的量化图像。事实上，量化的颜色和受损的结构会转移神经网络的注意力，

这可能会导致识别失败。例如，在图 1 的第二行中，在原始图像上训练的神经网络发现头部和身体对于识别狗最为关键。当颜色空间逐渐缩小时，神经网络首先无法关注头部，然后是身体，从而导致识别失败。这项工作的研究内容为如何在极小的色彩空间下保留关键结构。这个问题与传统的颜色量化问题正交，因为它是以任务为中心的：神经网络识别精度是它的主要关注点，而人类视觉不是关注点。如图 2 所示，这项工作使用在原始图像上预训练的分类器来评估量化的图像。为了优化识别精度，作者设计了一种颜色量化方法 ColorCNN，它以端到端的方式学习如何构建图像。与仅依赖颜色值进行决策的传统颜色量化方法不同，ColorCNN 利用颜色、结构和语义来发现和保留关键结构。它通过计算颜色索引图并分配调色板值来量化图像。在图 2 中，由 Color-CNN 量化的图像使分类器能够成功地聚焦于猫虎斑猫和前肢。在这个例子中，注意信息区域会导致正确识别。作者证明了 ColorCNN 量化对分类任务的有效性。仅使用几种颜色，该方法大大优于传统方法。

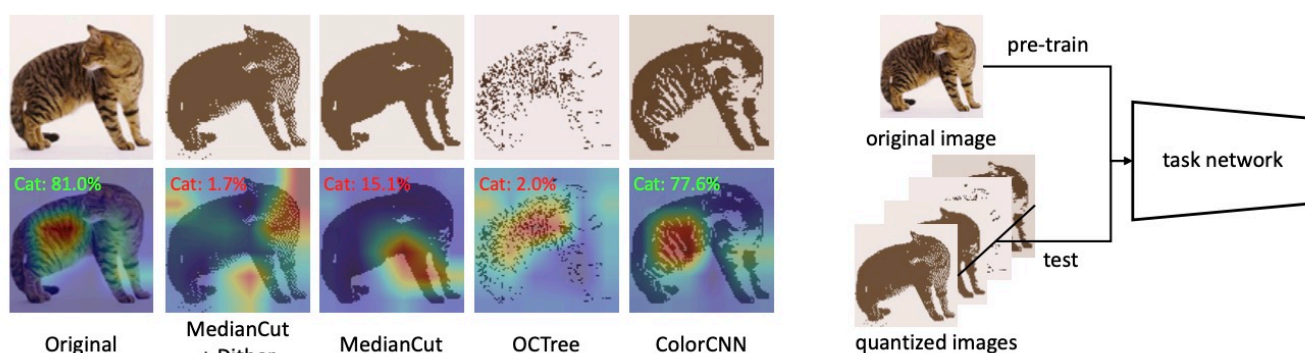


图 2: 左: 不同方法颜色量化结果; 右: 评估方式。左侧每列分别为原图, 1-bit 量化结果 (三种传统方法 MedianCut+Dither, MedianCut, OCTree, 文章提出的 ColorCNN 方法)。

2 相关工作

2.1 颜色量化

颜色量化^[2-6]对颜色进行聚类以减少颜色空间，同时保持视觉相似性。Heckbert 等人^[1]提出了常用的 MedianCut 方法。后来，Gervautz 等人^[7]设计了另一种常用的颜色量化方法: OCTree。此外，Dithering 通过添加噪声模式来消除视觉伪影，也作为一个可选步骤进行研究。颜色量化图像可以表示为颜色索引表，并用 PNG^[8] 编码。

2.2 以人类为中心的压缩

许多图像压缩方法都是为人类设计的。这些方法分为两类，无损压缩，例如 PNG^[8]，和有损压缩，例如 JPEG^[9]。最近，深度学习方法也被引入到图像压缩问题中，包括 RNN^[10] 和 CNN^[11]。Balle 等人^[12]提出了用于图像压缩的广义除法归一化 (GDN)。Agustsson 等人^[13]在生成对抗网络 (GAN) 中提出了一种用于低比特率压缩的多尺度鉴别器。在^[14-15]中，研究人员应用图像压缩方法来抵御对抗性攻击。

2.3 以任务为中心的压缩

传统或基于深度学习的上述提出到的图像压缩方法都是以人为中心的。Liu 等人^[16]指出对于分割，以人为中心的图片压缩不是 3D 医学图像的最佳选择。对于 2D 地图数据和 3D 场景模型，以任务为中心的压缩方法被设计用于定位^[17-18]。

2.4 神经网络对压缩数据的识别

一些研究人员使用压缩数据处理某些任务。例如，用压缩视频解决动作识别问题^[19-21]。Wang 等人^[22] 通过利用压缩数据加速视频对象识别。

3 本文方法

3.1 动机

在极小的颜色空间中，作者发现传统的颜色量化方法无法保留关键结构。这是因为这些方法通常采用仅颜色值的方法来聚类颜色，完全忽略结构。然而，正如^[23] 所暗示的那样，具有语义意义的关键形状和纹理在神经网络识别中起着重要作用。事实上，图 1 和图 2 中的结构没有很好地保存时，作者看到神经网络的关注点发生了偏移，进一步导致识别失败。受此失败的启发，文章将进一步研究如何在极小的颜色空间中有效地保留结构。

3.2 本文方法概述

为了提高神经网络在颜色量化图像上的性能，作者提出了 ColorCNN，可以端到端地学习如何进行颜色量化。如图 2，文章提出的 ColorCNN 有效保留了神经网络在猫的条纹和前爪的激活值，进而保证了压缩图像的正确识别。

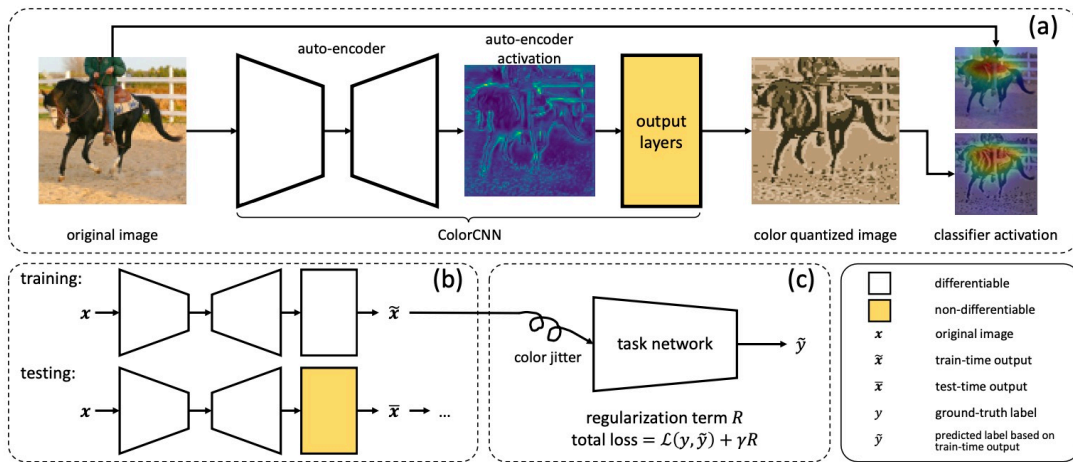


图 3: 本文所提出的方法概览。

本文提出的方法如图 3 所示。本文的目标是用很少的颜色构建图像，且预训练的分类器对颜色量化图像具有尽可能高的分类准确率。我们可以定义一个分类网络 $f_{\theta}(\cdot)$ ，其参数为 θ 。我们将 ColorCNN 的前向传播表示为一个函数 $\tilde{x} = g_{\psi}(x)$ ，参数为 ψ 。则我们的目标可以表示为：

$$\psi^* = \arg \min_{\psi} \sum_{(x,y \in D)} L(y, f_{\theta^*}(g_{\psi}(x))) + \gamma R, \quad (1)$$

其中， ψ^* 表示 ColorCNN 的优化参数， θ^* 为给定的预训练分类器的参数。 R 是正则化项， γ 表示其权重。

为了进行颜色量化，ColorCNN 输出一张颜色索引图和一个调色板。为了去除网络中的不可导部分（argmax 生成颜色序号图、解码时的查表），在训练过程中将它们都进行了替代，如图 4 所示。

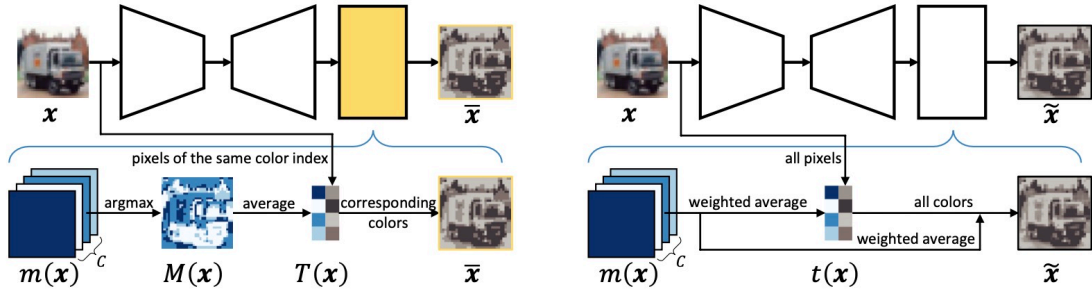


图 4: 测试和训练网络流程图对比。左：测试网络。右：训练网络。

可以看到，由于在解码时使用加权平均代替查表过程，故训练时 ColorCNN 会输出更多颜色。此外，在训练过程中，作者还引入了一个鼓励 ColorCNN 选择尽量多的颜色的约束：

$$R = \log_2 C \times (1 - \frac{1}{C} \times \sum_c \max_{(u,v)} [m(x)]_{u,v}), \quad (2)$$

我们取总和的负值，因为希望最小化这个正则化项 R 。我们还将正则化项偏移 1 以使其大于 0。

4 复现细节

4.1 与已有开源代码对比

原文章在 github 上开源了代码。本次复现工作对原文章的实验进行了复现：比较通过 ColorCNN 和传统方法得到的颜色量化图片在分类网络下的准确率。在复现中，数据集采用了 Tiny-ImageNet-200(tiny200), 分类网络采用了 AlexNet。具体步骤如下：a) 使用训练集从零开始训练 AlexNet 分类网络。b) 固定住分类网络的参数，训练 ColorCNN。在测试中得到通过 ColorCNN 进行颜色量化的图片在分类网络的准确率。c) 使用传统颜色量化方法 MedianCut, OCTree, MedianCut+Dither 得到颜色量化图片，将其输入到分类网络测试准确率。

4.2 创新点

本次复现工作的创新点如下：

- 研究分类任务准确率能否进一步提升。具体步骤如下：a) 在得到训练好的 ColorCNN 以后，输入数据集到 ColorCNN, 得到颜色量化的图片。b) 使用这些颜色量化后的图片对预训练的分类网络 fine-tune，测试其在分类网络上准确率。
- 研究颜色量化对 BPG^[24] 压缩的影响。具体步骤如下：a) 输入原始数据集到训练好的 ColorCNN，得到颜色量化后的图片。b) 使用 BPG 对颜色量化图片进行压缩，通过设置不同的 QP 得到不同码率的图片。c) 将压缩后的不同码率的图片输入到分类网络，测试分类准确率。
- 研究能否提升 BPG 压缩后的颜色量化图片在分类器的准确率。具体步骤如下：a) 使用此前已经得到 BPG 压缩后的颜色量化图片对预训练的分类网络 fine-tune，测试其在分类网络上的准确率。

5 实验结果分析

1. 复现原文实验结果。如图 5 所示，相比于传统颜色量化方法 MedianCut, OCTree, MedianCut+Dither, 使用 ColorCNN 得到的颜色量化图片能在分类任务上获得更高的准确率。

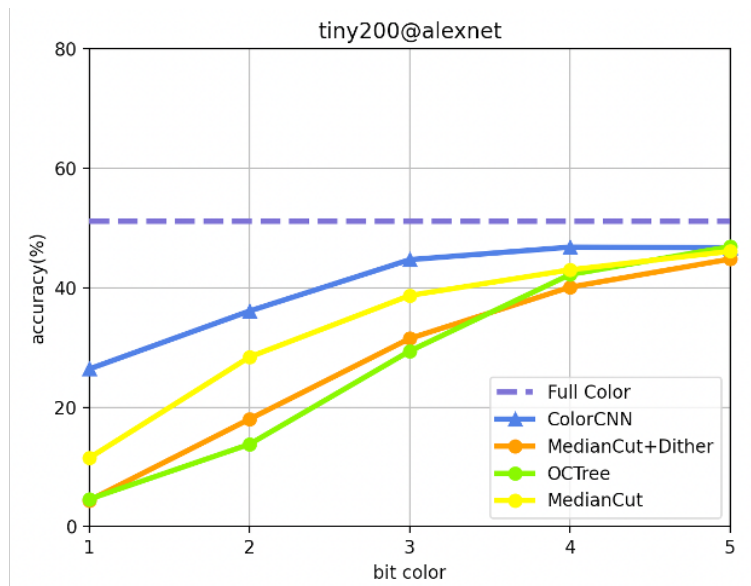


图 5: 原文实验复现结果。ColorCNN 相比传统颜色量化方法在分类任务上有更高的准确率。

2. 探究分类任务准确率能否进一步提升。如图 6所示，使用 ColorCNN 得到的颜色量化图片对分类器进行 fine-tune，可以进一步提高分类任务的准确率。

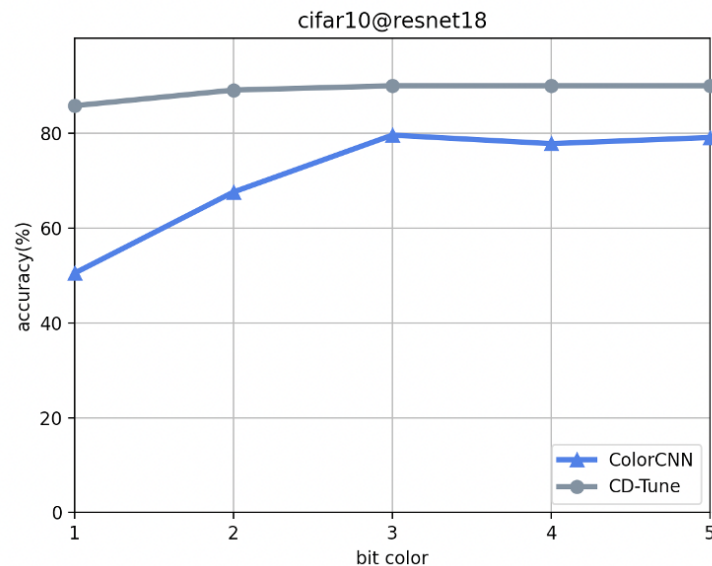


图 6: 对分类器 fine-tune 的准确率和原始准确率的比较。

3. 研究颜色量化对 BPG^[24] 压缩的影响。如图 7所示，相比于对原图进行压缩，压缩 1 bit 颜色空间的图片能提高压缩率，但是在压缩其它颜色空间的图片时，压缩率反而减小了。

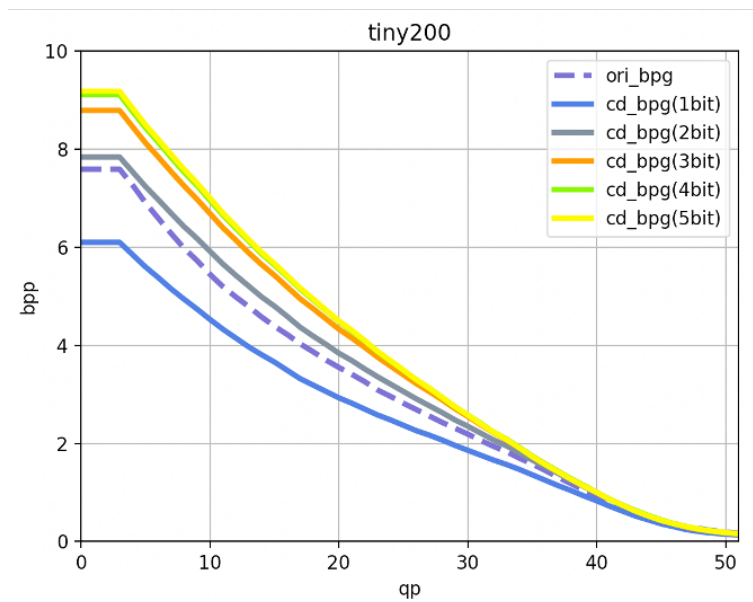


图 7: 使用 BPG 压缩 (QP 为 0-51) 颜色蒸馏后的图片 (颜色空间为 1bit-5bit) 以及压缩原图的 bpp (bits per pixel) 比较。

4. 研究能否提升 BPG 压缩后的颜色量化图片在分类器的准确率。使用 BPG 压缩后的颜色量化图片对预训练的分类网络进行 fine-tune, 在分类网络上测试的准确率如图 8 所示。整体而言, 相比原方法分类准确率有一定的提升。

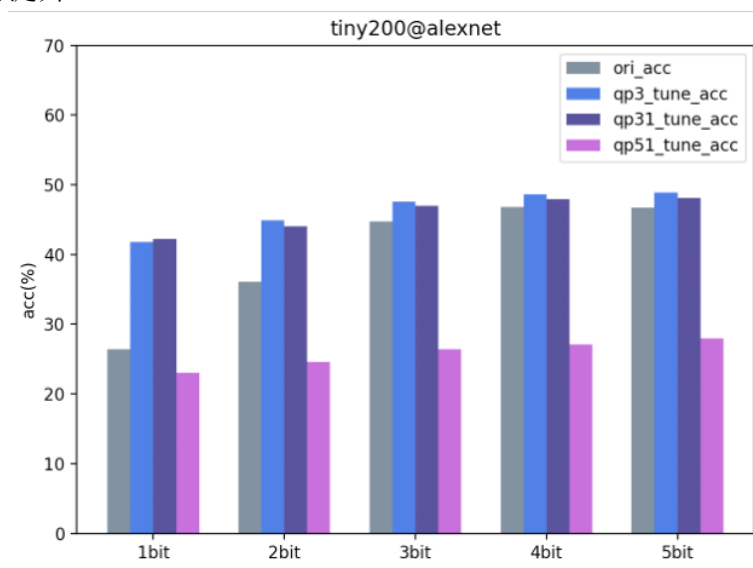


图 8: 使用 bpg 压缩的颜色量化图片在分类网络中的准确率 (颜色空间为 1bit-5bit, QP 取 3, 31, 51 三个点)。

其中在 1bit 和 2bit 颜色空间下, 准确率提升明显, 故在这两个颜色空间下选取了更多码率点来做验证实验, 结果如图 9 所示。

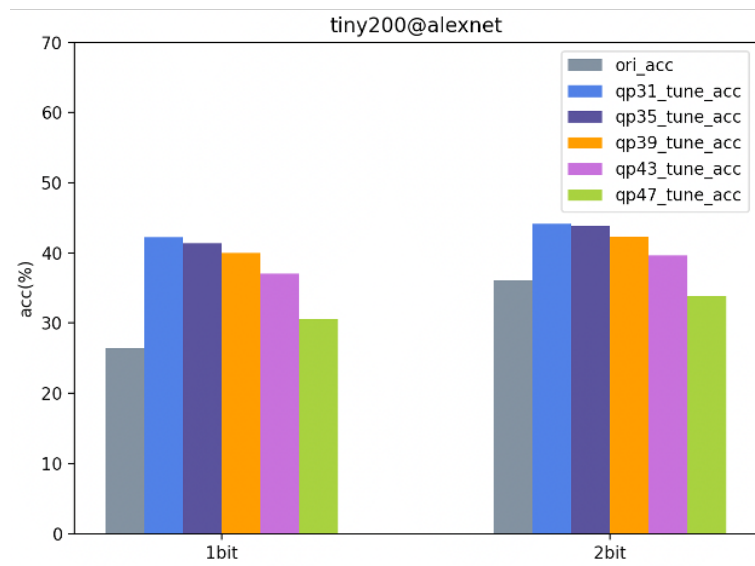


图 9: 使用 bpg 压缩的颜色量化图片在分类网络中的准确率（颜色空间为 1bit 和 2bit, QP 取 31, 35, 39, 43, 47 五个点）。

6 总结与展望

该文章研究了如何在极小的颜色空间下保持信息结构。传统的颜色量化方法往往会丢失颜色和结构,使其输出无法被神经网络理解。为了保持量化图像中的关键结构以便正确识别它们,文章提出了 ColorCNN 颜色量化网络,即使在极端条件下,ColorCNN 也能有效地识别和保留信息结构。文章为 ColorCNN 设计了一种端到端的训练方法,以最大限度地提高量化图像在神经网络任务上的性能,并通过验证在不同数据集上的分类任务的准确性来证明了方法的有效性。在复现的过程中,验证了颜色量化对 BPG 压缩的影响,以及分类任务准确率进一步提升的可能性,未来可在此方向上做进一步的研究。

参考文献

- [1] HECKBERT P. Color image quantization for frame buffer display[J]. ACM Siggraph Computer Graphics, 1982, 16(3): 297-307.
- [2] ORCHARD M T, BOUMAN C A, et al. Color quantization of images[J]. IEEE transactions on signal processing, 1991, 39(12): 2677-2690.
- [3] DENG Y, KENNEY C, MOORE M S, et al. Peer group filtering and perceptual color image quantization [C]//1999 IEEE International Symposium on Circuits and Systems (ISCAS): vol. 4. 1999: 21-24.
- [4] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(11): 2274-2282.
- [5] DENG Y, MANJUNATH B S. Unsupervised segmentation of color-texture regions in images and video [J]. IEEE transactions on pattern analysis and machine intelligence, 2001, 23(8): 800-810.
- [6] WU X. Color quantization by dynamic programming and principal analysis[J]. ACM Transactions on Graphics (TOG), 1992, 11(4): 348-372.

- [7] GERVAUTZ M, PURGATHOFER W. A simple method for color quantization: Octree quantization[C] //New Trends in Computer Graphics: Proceedings of CG International' 88. 1988: 219-231.
- [8] BOUTELL T. Png (portable network graphics) specification version 1.0[R]. 1997.
- [9] WALLACE G K. The JPEG still picture compression standard[J]. IEEE transactions on consumer electronics, 1992, 38(1): xviii-xxxiv.
- [10] VAN DEN OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[C] //International conference on machine learning. 2016: 1747-1756.
- [11] MENTZER F, AGUSTSSON E, TSCHANNEN M, et al. Practical full resolution learned lossless image compression[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 10629-10638.
- [12] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimized image compression[J]. arXiv preprint arXiv:1611.01704, 2016.
- [13] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 221-231.
- [14] LIU Z, LIU Q, LIU T, et al. Feature distillation: Dnn-oriented jpeg compression against adversarial examples[C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 860-868.
- [15] JIA X, WEI X, CAO X, et al. Comdefend: An efficient image compression model to defend adversarial examples[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 6084-6092.
- [16] LIU Z, XU X, LIU T, et al. Machine vision guided 3d medical image compression for efficient transmission and accurate segmentation in the clouds[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 12687-12696.
- [17] WEI X, BÂRSAN I A, WANG S, et al. Learning to localize through compressed binary maps[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10316-10324.
- [18] CAMPOSECO F, COHEN A, POLLEFEYS M, et al. Hybrid scene compression for visual localization [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7653-7662.
- [19] WU C Y, ZAHEER M, HU H, et al. Compressed video action recognition[C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6026-6035.
- [20] ZHANG B, WANG L, WANG Z, et al. Real-time action recognition with enhanced motion vector CNNs [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2718-2726.

- [21] SHOU Z, LIN X, KALANTIDIS Y, et al. Dmc-net: Generating discriminative motion cues for fast compressed video action recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 1268-1277.
- [22] WANG S, LU H, DENG Z. Fast object detection in compressed video[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 7104-7113.
- [23] GEIRHOS R, RUBISCH P, MICHAELIS C, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness[J]. arXiv preprint arXiv:1811.12231, 2018.
- [24] BPG image format[Z]. <https://bellard.org/bpg/>. Accessed: 2023-02-09. 2014.