基于学习的跨模态图像压缩

许育瑜

摘要

多模态（即多传感器）数据广泛用于各种视觉任务，以获得更准确或更稳健的感知。然而，增加的数据模态给数据存储和传输带来了新的挑战。现有的数据压缩方法通常对每种模态采用单独的编解码器，而没有考虑不同模态之间的相关性。这项工作通过利用跨模态冗余为红外和彩色图像对提出了多模态压缩框架。具体来说，给定参考模态中的图像（例如，可见光图像），我们使用逐通道对齐模块基于仿射变换生成对齐特征。然后将对齐的特征用作压缩当前模态（例如，红外图像）中图像的上下文信息，并且以可忽略的成本无损压缩相应的仿射系数。此外，我们引入了基于 Transformer 的空间对齐模块，以利用不同模态的解码过程中中间特征之间的相关性。实验结果表明，我们提出的框架在 FLIR 数据集上优于传统的和基于学习的单模态压缩方法。

关键词：多模态压缩框架，通道对齐模块，空间对齐模块

1 引言

在一些实际的视觉应用（例如自动驾驶）中，来自不同模态的相机（例如可见光或红外成像相机）通常通过利用互补特性联合用于各种计算机视觉任务[1-3]。例如，可见光 (RGB) 摄像头通常可以提供连续的高分辨率彩色图像，但在极低光照场景下可能效果不佳，而这正是红外摄像头可以提供的帮助。同时，红外摄像机容易受到异常热源的干扰，但使用可见光摄像机可以弥补这一缺点。然而，这些多模态视觉分析方法，随着来自不同模态的更多图像被传输到解码器端进行视觉分析，将增加存储和传输成本。因此，如何设计一种高效的多模态视觉数据压缩方法是一个新的、具有挑战性的研究问题。

在过去的几十年中，许多传统的和基于学习的压缩方法被提出用于图像或视频压缩[4-7]。然而，现有的大多数工作都集中在单模态图像压缩上，而没有考虑不同模态之间的相关性。由于来自不同模态的图像之间的强相关性，我们无法使用现有的单模态压缩方法来充分利用压缩冗余。最相关的研究课题之一是立体图像压缩，其中通过使用各种视图对齐方法来利用交叉视图冗余。然而，与具有相似分布的立体图像相比，不同模态图像的强度可能有很大差异（见图 1）。因此，常用的对齐技术，如基于块的运动/视差估计[8]或单应变换[9]对于多模态压缩来说不够可行。此外，考虑到红外图像和可见光图像对等多模态数据代表不同视角下的同一场景，大多数现有估计方法对像素运动/视差信息的压缩将消耗大量比特，因此，开发一个新的多模态数据压缩框架非常重要。

在本文中，我们通过利用特征空间中的跨模态冗余，为红外和可见图像对提出了一种基于学习的多模态压缩框架。考虑到不同模态的显式对齐非常困难，并且估计的运动/视差信息也需要大量的传输比特率，我们使用有效的仿射变换和注意机制分别实现通道和空间特征对齐。具体地，以红外图像的压缩过程为例，根据解码后的可见光图像（即 RGB 图像）和原始红外光图像提取的特征，估计仿射变换系数，并以可忽略的带宽成本将其传输给解码端。然后我们实现了基于仿射变换的逐通道特征对齐，并将来自可见模态的相应变换特征用作压缩红外图像的条件上下文。此外，我们通过空间对齐模块在解码过程中利用来自不同模态的中间特征的相关性。我们的模块集成到可见图像解码器中，并

将在空间上扭曲参考模态的中间特征以生成对齐特征，用于进一步减少跨模态冗余。我们框架的贡献总结如下：

• 我们提出了一个基于学习的框架，通过利用跨模态冗余来压缩来自不同模态的图像对。

• 我们的框架引入了通道对齐和空间对齐模块，以有效利用特征空间中不同模态之间的相关性。

• 实验结果表明，本文提出的方法在 FLIR 数据集上比单模态图像方法实现了更好的压缩性能。

2 相关工作

2.1 图像和视频压缩

在过去的几十年里，几个有代表性的压缩标准[4-5,10]被提出并广泛应用于许多实际应用中。最近，基于学习的图像和视频压缩方法引起了越来越多的关注[11-13]并显示出可比性甚至比最新的图像或视频压缩标准更好的性能[5,14]。尽管将这些方法扩展到红外图像或视频压缩[11,15]是可行的，但现有标准只能减少单一模态的冗余，而不能利用跨模态信息。考虑到存储和传输多模态数据（如深度图、红外图像或光流图）的需求不断增加，有必要为多模态数据提出一种新的压缩框架。

2.2 立体图像和视频压缩

立体图像压缩旨在压缩一对来自不同视图的图像。为了利用这种视图间冗余，基于传统的单视图图像/视频方法提出了几种多视图图像/视频压缩标准，如 MV-HEVC[8]或 MVC[16]。除了现有的帧间补偿之外，这些方法还使用基于视差的运动补偿[17]来提高压缩性能。

最近的工作还尝试使用深度神经网络进行立体图像压缩[9,18]。文献[18]引入了参数跳跃函数以利用参考视图中的视差补偿功能。在[9]中，估计单应性矩阵将左视图图像扭曲到右视图图像，从而减少视图冗余。然而，这些学习到的立体图像压缩方法仍然用于由位置略有不同的立体相机记录的单模态图像。使用不同的相机捕获多模态数据，例如可见光和红外配对图像。这些图像的内部特征差异很大，单应性变换等现有技术不可行。因此，有必要开发一种多模态图像压缩框架。

2.3 多模态数据压缩

多模态或多传感器信息广泛用于各种计算机视觉任务[1,3,19]尤其是 3D 视觉任务。例如，文献[3]利用图像和点云信息来提高 3D 对象检测的准确性。文献[20]从不同的模态中提取特征并融合这些特征以进行对象跟踪。

近年来，一些多模态数据压缩方法被提出[21-23]。然而，这些方法是基于手工制作的编解码器，主要是为多视图图像和深度图像的联合压缩，或者医学图像和信号的联合压缩而设计的。因此，对可见光-红外对压缩的研究尚属空白。

3 本文方法

3.1 本文方法概述

我们的多模态压缩方法的总体架构如图 1所示。这里我们使用重建的可见图像 x^v 作为跨模态参考来提高红外图像 x^i 的压缩性能。

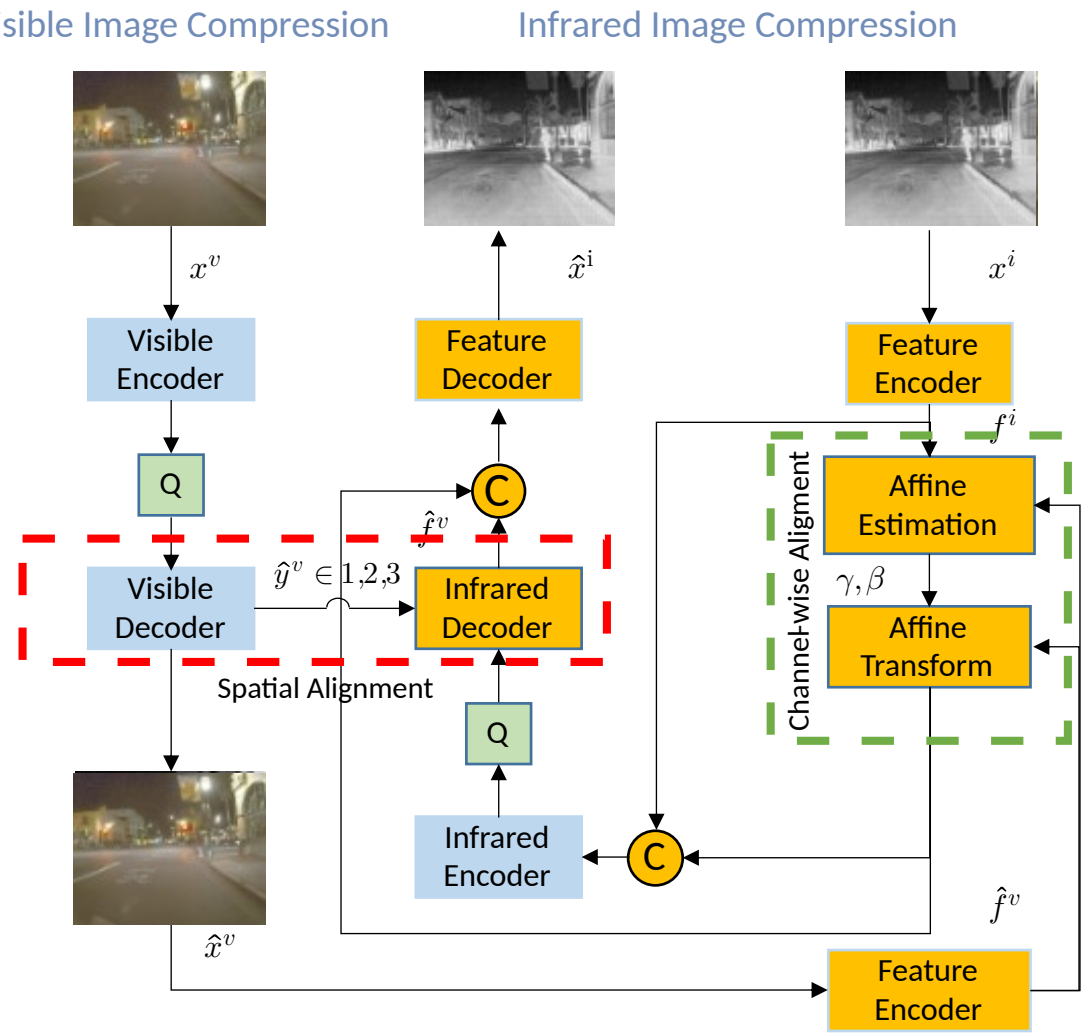


图 1: 方法示意图（黄色模块为复现部分）

如图 2所示，我们首先使用现有的图像压缩方法[24]对可见图像 x^v 进行压缩。然后，红外图像 x^i 和重建的可见光图像 x^v 的特征由特征编码器模块提取，该模块使用多个卷积层实现。基于提取的特征，引入通道特征对齐模块来计算通道仿射变换系数 β 和 γ，以将红外模态的特征与可见光模态对齐。在我们的框架中，β 和 γ 被无损地传输到解码器端。之后，将对齐的特征 $\bar{f}^i$ 作为上下文信息输入到红外图像编码器网络中。在这里，我们遵循文献[24]中的网络设计来实现图像编解码器。最后，红外图像解码器的输出 $\hat{f}^i$ 与对齐的特征 $\bar{f}^i$ 连接起来通过特征解码器产生重构帧 $\hat{x}^i$。

考虑到不同模态特征之间的空间相关性在通道对齐模块中没有得到充分利用，我们通过可见解码器中的空间对齐模块进一步利用不同模态的中间特征之间的相关性。如图 3所示，$\hat{y}_j^i$ 和 $\hat{y}_j^v$ 分别表示红外和可见解码器中第 j 个反卷积层的输出。我们的空间特征对齐模块使用基于 Transformer 的机制将中间特征从可见光模态空间扭曲到红外模态，并且在解码过程中使用扭曲的特征。第 3.3 节提供了更多详细信息。由于篇幅有限，详细的补充材料参见文献[25]的补充材料中提供的特征编码器/解码器和可见/红外编解码器（编码器和解码器）的网络架构。

3.2 通道对齐模块

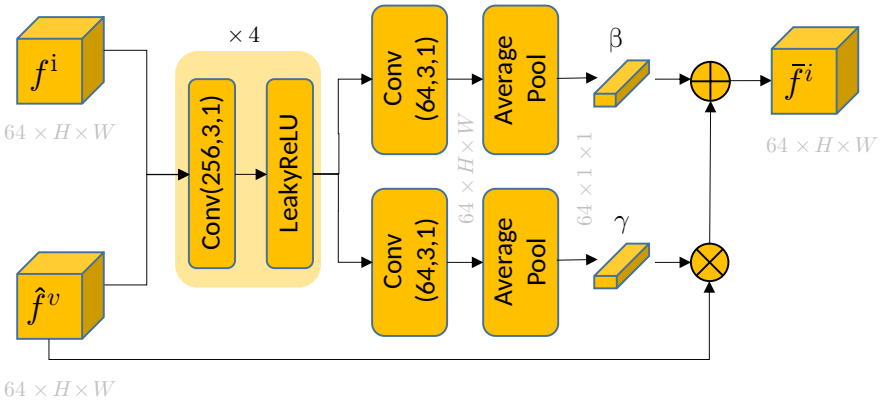


图 2: 通道对齐模块（黄色模块为复现部分）

在我们提出的框架中，我们使用特征空间中的逐通道对齐来减少重建的可见图像 $\hat{x}^v$ 和红外图像 x_i 之间的冗余。我们的通道对齐模块的网络架构如图 2所示。给定从可见图像和重建的红外图像中提取的特征 $\hat{f}^v$ 和 f^i，我们将它们提供给几个卷积层。之后，我们使用空间平均池生成仿射变换系数 $\gamma, \beta \in R^{64\times1\times1}$。然后将来自解码可见图像的特征 $\hat{f}^v$ 与红外模态对齐，如下所示，

$$\bar{f}^i = \gamma \times \hat{y}^v + \beta \tag{1}$$

其中 $\times$ 和 $+$ 分别代表通道乘法和加法。$\bar{f}^i$ 是对齐的特征图。在编码器端，将对齐特征 $\bar{f}^i$ 和原始特征 f_i 连接起来作为后续编码器网络的输入。

在解码器端，接收到的仿射变换系数 β 和 γ 用于产生对齐的特征 $\bar{f}_i$，它将与解码器的输出连接起来，通过特征解码器获得最终的重构帧 $\hat{x}^v$。考虑到这些系数是紧凑的，我们不进行任何压缩，它们以可忽略的成本无损地发送到解码器端。

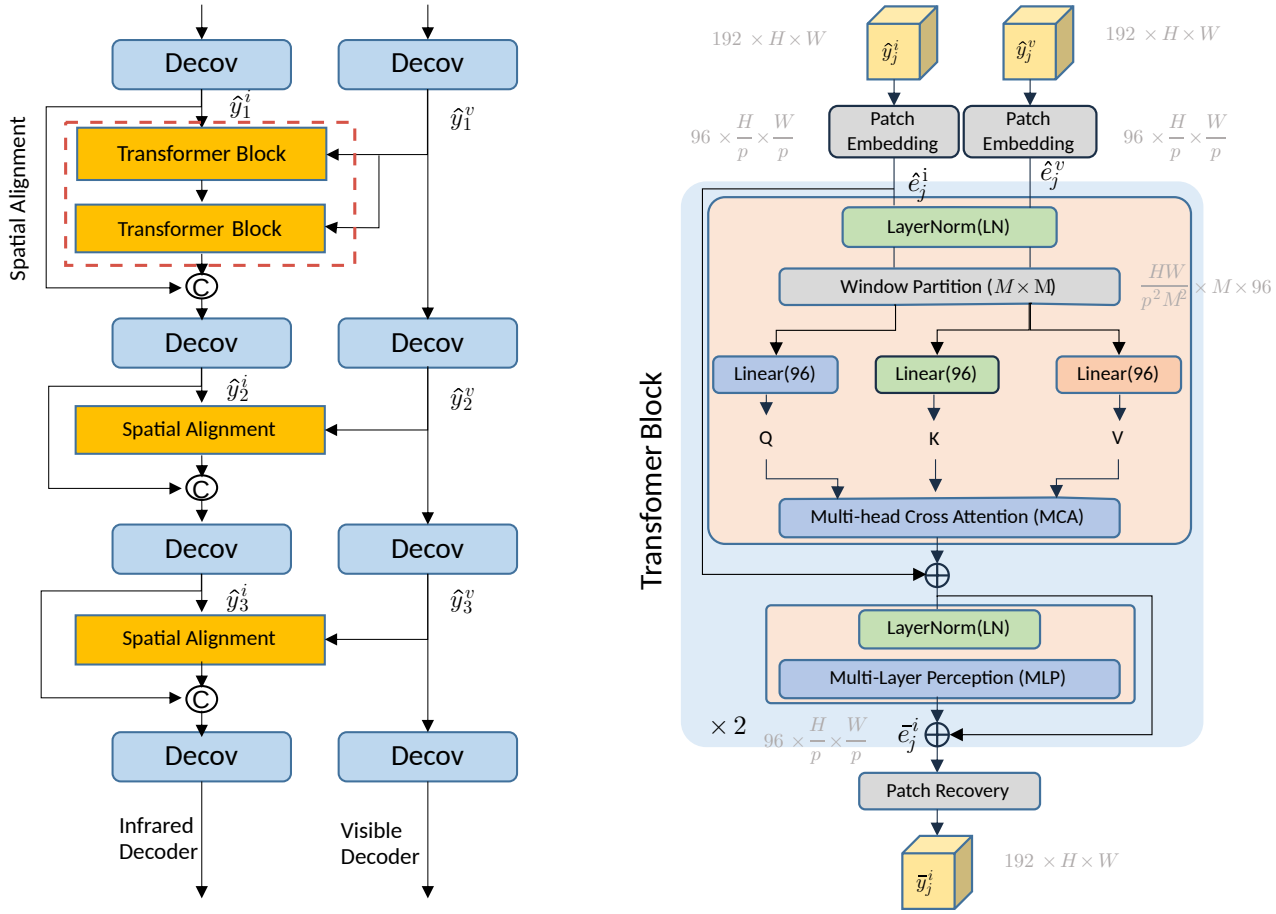
3.3 空间对齐模块

由于通道对齐模块仅通过逐通道变换利用跨模态冗余，因此未充分利用不同模态中特征之间的空间相关性。我们的可见解码器使用空间特征对齐模块，根据这两个特征之间的相似性，将特征从可见光模态空间扭曲到红外模态。空间对齐模块的整个网络架构如图 3所示。受 Swin-Transformer[26]的启发，我们使用基于 Transformer 的机制在解码过程中利用可见光图像 x^v 和红外图像 x^i 的中间特征之间的相关性。

具体来说，让 $\hat{y}_j^v \in R^{192\times H\times W}$ 和 $\hat{y}_j^i \in R^{192\times H\times W}$ 分别表示图 3(a)中 x^v 和 x^i 的解码器网络的第 j 个反卷积层的输出。我们首先使用卷积层执行 $p \times p$ 块嵌入操作，并生成相应的嵌入 $\hat{e}_j^v \in R^{192\times H\times W}$ 和 $\hat{e}_j^i \in R^{192\times H\times W}$，其中 p 设置为 2。然后，将 $\hat{e}_j^v$ 和 $\hat{e}_j^i$ 送入 LayerNorm 和多头交叉注意力（MCA）模块，其中来自不同模态的特征用于计算注意力矩阵，可见光嵌入特征 $\hat{e}_j^v$ 被扭曲以生成对应的对齐特征 $\tilde{e}_j^v$。之后，我们使用 LayerNorm 和 MLP 网络进一步增强特征变换[26]。此外，添加了残差连接以帮助训练过程，这个基于 Transformer 的块的公式如下，

$$\begin{aligned} \hat{e}_j^i &= MCA(LN(\hat{e}_j^i), LN(\hat{e}_j^i)) + \hat{e}_j^v \\ \bar{e}_j^i &= MLP(LN(\bar{e}_j^i)) + \bar{e}_j^i \end{aligned} \tag{2}$$

在我们的实现中，我们使用两个 Transformer 块并且嵌入特征会在输入第二个 Transformer 块之前移动。最后，使用反卷积层将生成的嵌入 $\bar{e}_j^i$ 恢复为中间特征，这是块嵌入的逆过程。我们在解码器端使用三



(a) 与解码器的连接结构　　(b) 具体结构

图 3: 空间对齐模块（黄色模块为复现部分）

个空间特征对齐模块，我们模块的输出将被馈送到下一个反卷积层，如图 3(a)所示。

多头交叉注意力。多头交叉注意模块将从可见光图像生成对齐嵌入特征。具体来说，将输入的嵌入特征划分为形状为 $\frac{HW}{p^2M^2} \times M^2 \times 96$ 的不重叠的 $M \times M$ 窗口，其中 $\frac{HW}{p^2M^2}$ 表示窗口的数量，M 设置为 4。然后，模块计算红外模态和可见模态的窗口之间局部注意力。以可见光和红外图像嵌入中的第 n 个局部窗口 $\hat{e}_j^v$ 和 $\hat{e}_j^i \in R^{M^2 \times 96}$ 为例，对应的查询、键、值矩阵 Q、K、$V \in R^{M^2 \times 96/h \times h}$ 计算如下，

$$Q = \hat{e}_j^i(n)P_Q, K = \hat{e}_j^v P_K, V = \hat{e}_j^v(n)P_V \tag{3}$$

其中 h 是多头注意力中的头数，设置为 3，P_Q，P_K 和 P_V 是跨窗口共享的空间特征对齐模块中的投影矩阵。然后将从可见光图像特征生成的值 V 与红外特征对齐，如下所示，

$$A = SoftMax(QK^T/\sqrt{d} + B)V \tag{4}$$

其中 B 是可学习的相对位置编码，$d = 96/h$ 是每个头中的通道数。A 是局部窗口 $\hat{e}_j^v(n)$ 的多头交叉注意力机制 (MCA) 输出，被认为是从可见光图像到红外图像的对齐嵌入结果。

3.4 损失函数

红外图像的压缩网络通过使用以下率失真损失函数进行优化，

$$\lambda D + R = \lambda d(x^i, \hat{x}^i) + H(\hat{y}^i) + H(\gamma) \tag{5}$$

其中，$\lambda d(x^i, \hat{x}^i)$ 表示输入图像 x^i 和重建图像 $\hat{x}^i$ 之间的失真。$H(\cdot)$ 表示用于对表示进行编码的位数。在我们的框架中，潜在表示 $\hat{y}^i$ 通过使用[24]中的熵模型进行编码，并且通道仿射变换系数 γ，β 以 Float 格式直接存储和传输，带宽成本可忽略不计。λ 是用于控制率失真权衡的超参数。

4 复现细节

4.1 与已有开源代码对比

编码器和解码器模块：该模块的实现主要参照文献[27]中对文献中[24]算法的实现。可见光图像的压缩是在 FLIR 数据集上使用文献[24]中模型进行训练，然后将解码器部分的每层上采样的中间结果特征输出到空间对齐模块中。文献[27]实现的编解码器模块仅仅支持 3 通道图像的压缩，本文将其修改成可以对多通道特征进行压缩的模块，然后将红外图像经过特征提取和通道对齐之后获得的特征进行编码和解码。

通道对齐模块、特征提取模块：由于这两个模块没有开源代码，本文参照复现文献[25]以及其补充材料中对网络结构的描述，使用 pytorch 编写出相应模块的代码来分别实现跨模态特征的通道对齐，以及对两个模态的特征提取。

空间对齐模块：该模块的实现主要参照文献[26]中的 swin-transformer 模块实现。在文献[26]中，swin-transformer 模块快速高效地提取图像本身的多尺度特征；本文则对 swin-transformer 模块修改成能够将两个不同模态图像的多尺度特征在空间上进行对齐，从而减少不同模态图像之间的冗余。

4.2 实验环境搭建

FLIR Thermal 数据集[28] 包含超过 10K 对的 8 位红外图像和 24 位可见光图像，包括白天和夜晚场景中的人、车辆、自行车和其他物体。红外图像的分辨率为 640×512，而对应的可见光图像分辨率从 720×480 到 2048×1536 不等。我们在实验中将每个可见图像的大小调整为 1280×1024。默认的 FLIR 训练数据集用作我们的训练数据集，并从 FLIR 验证集中随机选择 20 个可见光-红外图像对作为测试数据集。

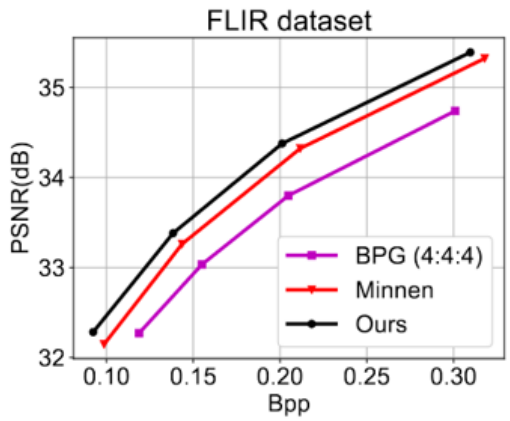
评估指标 bpp（每像素位数）衡量压缩过程中的平均位数消耗，PSNR 来测量重建图像与原始红外图像之间的失真。

实现细节当我们使用可见光图像作为参考对红外图像进行编码时，我们首先训练可见光图像压缩网络，然后通过冻结可见光图像压缩来优化可见数据压缩网络。这些网络基于具有 CUDA 支持的 PyTorch 实现，并在 V100 GPU 卡上进行训练。具体来说，对于多模态图像压缩，我们设置不同的 λ 值（λ= 256、512、1024、2048、4096）并通过将初始学习率 β1 和 β2 分别设置为 1e-4、0.9、0.999 来使用 Adam 优化器。当损失变得稳定时，学习率在 1.8M 步后降低到 1e-5。mini-batch 大小设置为 4。训练阶段大约需要 8 天。

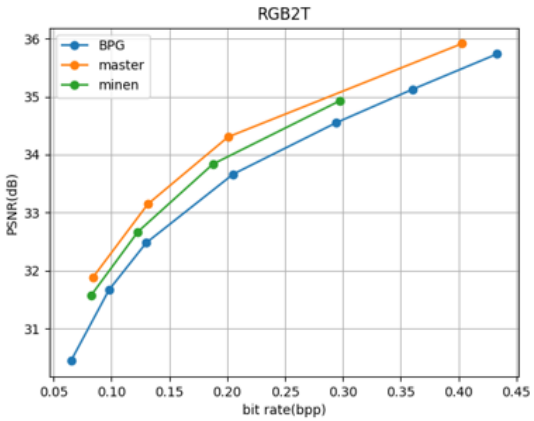
5 实验结果分析

为了证明我们方法的有效性,我们将我们的方法与传统的单模态图像压缩方法 BPG[29]和文献[24]提出的学习图像压缩方法在 FLIR 测试数据集上进行了比较。此外，为了公平比较，我们的模型和基线方法[24]都使用基于 MSE(即 PSNR）指标的相同多模态数据进行了优化。

图 4显示了 FILR 数据集上可见图像压缩的不同压缩方法的率失真曲线, 本文复现的结果与原论文结果相近。与单独优化的单模态压缩方法[24]相比，我们使用可见光图像作为参考的方法可以在 FLIR 数据集上将压缩性能提高 0.31dB 以上。此外，我们的方法还实现了比传统图像压缩方法 BPG 更好的压缩性能。

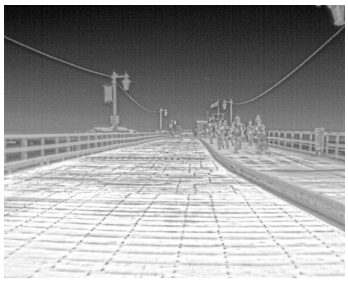


(a) 论文结果

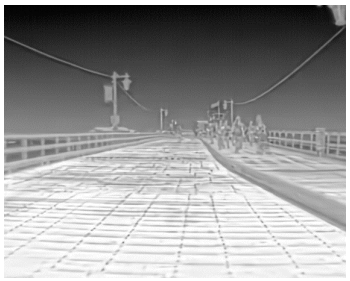


(b) 复现结果

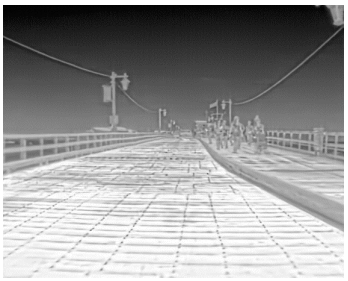
图 4: 不同算法对红外图像压缩的 PSNR 结果对比



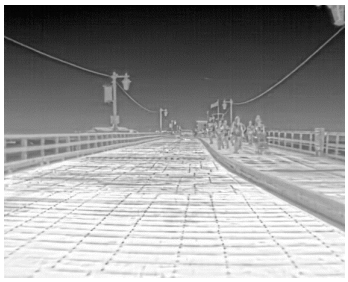
(a) raw



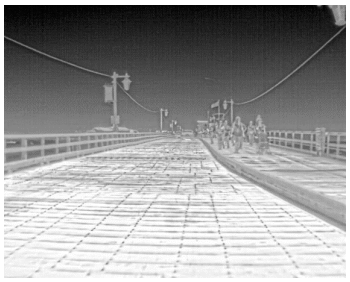
(b) $\lambda = 256$: 0.11bpp, 29.73dB



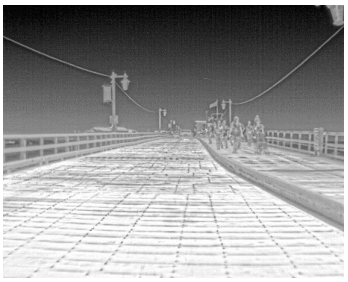
(c) $\lambda = 512$: 0.16bpp, 30.75dB



(d) $\lambda = 1024$: 0.24bpp, 31.93dB



(e) $\lambda = 2048$: 0.57bpp, 34.28dB



(f) $\lambda = 4096$: 1.12bpp, 38.04dB

图 5: 本论文算法在不同不同码率下对红外图像压缩示例

如图 5(a)所示，原始的红外图像能够看见远处的电线，以及天上的流星。本文跨模态压缩算法在最低码率下（如图 5(b)、5(c)）虽然不能看见远方的电线和天空的流星，仍然能够很好地还原红外图像中的大部分信息。此外，低码率的压缩图像整体比未压缩图像变得更加平滑。随着码率的不断增加，图 5(d)、5(e)、5(f)中的图像远处的电线和流星越来越清晰，而图 5(f)甚至将原始图像中的噪声信息也还原了出来。通过图 5可以看出，本文所提出的跨模态图像压缩算法在不同码率下都能够对红外图像有良好的压缩效果。

6 总结与展望

在这项工作中，我们提出了一种用于可见光和红外图像对的多模态压缩框架。为了利用互补信息，我们引入了通道方式和空间特征对齐模块。在 FLIR 数据集上的实验结果证明了我们的多模态图像压缩方法的有效性。此外，我们的框架还可以扩展到其他彼此接近的多模态数据，以及多模态的视频压缩。将来，我们将研究新的压缩方法来压缩更具挑战性的多模态数据。

参考文献

[1] DENG X, DRAGOTTI P L. Deep Convolutional Neural Network for Multi-Modal Image Restoration and Fusion[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3333-3348. http://dx.doi.org/10.1109/TPAMI.2020.2984244.

[2] CHAVEZ-GARCIA R O, AYCARD O. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking[J/OL]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(2): 525-534. http://dx.doi.org/10.1109/TITS.2015.2479925.

[3] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection[C/OL]//: vol. 2019-June. Long Beach, CA, United states, 2019: 7337-7345. http://dx.doi.org/10.1109/CVPR.2019.00752.

[4] SCHWARZ H, MARPE D, WIEGAND T. Overview of the scalable video coding extension of the H.264/AVC standard[J]. IEEE Transactions on circuits and systems for video technology, 2007, 17(9): 1103-1120.

[5] BROSS B, WANG Y K, YE Y, et al. Overview of the versatile video coding (VVC) standard and its applications[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3736-3764.

[6] BALLÉ J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior[J]. arXiv preprint arXiv:1802.01436, 2018.

[7] AGUSTSSON E, MENTZER F, TSCHANNEN M, et al. Soft-to-hard vector quantization for end-to-end learning compressible representations[J]. Advances in neural information processing systems, 2017, 30.

[8] HANNUKSELA M M, YAN Y, HUANG X, et al. Overview of the multiview high efficiency video coding (MV-HEVC) standard[C]//2015 IEEE International Conference on Image Processing (ICIP). 2015: 2154-2158.

[9] DENG X, YANG W, YANG R, et al. Deep homography for efficient stereo image compression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 1492-1501.

[10] SKODRAS A, CHRISTOPOULOS C, EBRAHIMI T. The JPEG 2000 still image compression standard [J]. IEEE Signal processing magazine, 2001, 18(5): 36-58.

[11] FIDALI M, JAMROZIK W. Compression of high dynamic infrared image using auto aggregation algorithm[J]. Measurement Automation Monitoring, 2017.

[12] DJELOUAH A, CAMPOS J, SCHAUB-MEYER S, et al. Neural Inter-Frame Compression for Video Coding[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 6420-6428.

[13] CHENG Z, SUN H, TAKEUCHI M, et al. Learned Image Compression With Discretized Gaussian Mixture Likelihoods and Attention Modules[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 7936-7945.

[14] SULLIVAN G J, OHM J R, HAN W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on circuits and systems for video technology, 2012, 22(12): 1649-1668.

[15] LI J, FU Y, LI G, et al. Remote Sensing Image Compression in Visible/Near-Infrared Range Using Heterogeneous Compressive Sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11: 4932-4938.

[16] VETRO A, WIEGAND T, SULLIVAN G J. Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard[J]. Proceedings of the IEEE, 2011, 99: 626-642.

[17] SCHWARZ H, WIEGAND T. Inter-view prediction of motion data in multiview video coding[J]. 2012 Picture Coding Symposium, 2012: 101-104.

[18] LIU J, WANG S, URTASUN R. DSIC: Deep Stereo Image Compression[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3136-3145.

[19] GARCÍA R O C, AYCARD O. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17: 525-534.

[20] ZHANG W, ZHOU H, SUN S, et al. Robust Multi-Modality Multi-Object Tracking[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 2365-2374.

[21] BRAHIMI T, BOUBCHIR L, FOURNIER R, et al. An improved multimodal signal-image compression scheme with application to natural images and biomedical data[J]. Multimedia Tools and Applications, 2017, 76: 16783-16805.

[22] CHEN S, LIU Q, YANG Y. Adaptive Multi-Modality Residual Network for Compression Distorted Multi-View Depth Video Enhancement[J]. IEEE Access, 2020, 8: 97072-97081.

[23] VARADARAJAN K M, ZHOU K, VINCZE M. RGB and depth intra-frame Cross-Compression for low bandwidth 3D video[J]. Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012: 955-958.

[24] MINNEN D C, BALLÉ J, TODERICI G. Joint Autoregressive and Hierarchical Priors for Learned Image Compression[J]. ArXiv, 2018, abs/1809.02736.

[25] LU G, ZHONG T, GENG J, et al. Learning based Multi-modality Image and Video Compression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 6083-6092.

[26] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

[27] BÉGAINT J, RACAPÉ F, FELTMAN S, et al. CompressAI: a PyTorch library and evaluation platform for end-to-end compression research[J]. arXiv preprint arXiv:2011.03029, 2020.

[28] Flir thermal dataset[EB/OL]. https://www.flir.com/oem/%20adas/adas-dataset-form/.

[29] BELLARD F. bpg image format[EB/OL]. http://bellard.org/%20bpg/.