Optoelectronics 光电子, 2022, 12(4), 147-158
Published Online December 2022 in Hans. http://www.hanspub.org/journal/oe
https://doi.org/10.12677/oe.2022.124017

Hans汉斯

基于自适应卷积的视差细化网络

胡文辉，周泽豪，戚　桢

天津工业大学，天津

收稿日期：2022年11月16日；录用日期：2022年12月16日；发布日期：2022年12月28日

摘　要

为提高视差细化的精度，本文提出一种基于自适应卷积的视差细化与采样方法。利用细化假设为视差细化引入其它可用信息，在不同阶段附加不同的信息来增强细化假设。基于自适应传播方法构建局部代价卷，并将聚合操作从空间域转换至视差域，以缓解使用大卷积窗口带来的边界模糊问题，增强在无纹理或弱纹理区域的聚合效果。同时，使用自适应卷积从相似视差平面上更新视差，进而提高视差的精度。对于上采样过程，利用视差自适应采样克服双线性插值导致的精度下降问题。在SceneFlow和KITTI2015数据集上，对算法进行验证，实验结果表明，相比原始方法，本文算法在精度方面有了明显提升，特别是在KITTI2015数据集上，端点误差(EPE)和3像素错误率指标分别提升9.7%和12.5%。

关键词

机器视觉，立体匹配，双目视觉，自适应卷积

Parallax Refinement Network Based on Adaptive Convolution

Wenhui Hu, Zehao Zhou, Zhen Qi

Tiangong University, Tianjin

Received: Nov. 16th, 2022; accepted: Dec. 16th, 2022; published: Dec. 28th, 2022

Abstract

To improve the accuracy of disparity refinement, this paper proposes a disparity refinement and sampling method based on adaptive convolution. The refine hypothesis is used to introduce other available information for disparity refinement, and different information is attached at different stages to augment the refine hypothesis. Based on the adaptive propagation method, the local cost volume is constructed and the aggregation operation is converted from the spatial domain to the disparity domain to alleviate the boundary blur problem caused by using large convolution windows and augmented the aggregation effect in texture-free or weakly textured areas. At the same

文章引用: 胡文辉, 周泽豪, 戚桢. 基于自适应卷积的视差细化网络[J]. 光电子, 2022, 12(4): 147-158.
DOI: 10.12677/oe.2022.124017

time, adaptive convolution is used to update the disparity from similar disparity planes, which in turn improves the accuracy of disparity. For the upsampling process, disparity adaptive sampling is used to overcome the degradation of accuracy caused by bilinear interpolation. The algorithm is validated on the SceneFlow and KITTI2015 datasets, and the experimental results show that the algorithm in this paper has significantly improved in terms of accuracy compared with the original method, especially on the KITTI2015 dataset, the Endpoint Error (EPE) and 3-pixel error rate indicators have increased by 9.7% and 12.5%, respectively.

Keywords

Machine Vision, Stereo Matching, Binocular Vision, Adaptive Convolution

1. 引言

大多数基于卷积神经网络的端到端立体匹配网络都是模仿传统的立体匹配流程[1]，即由特征提取、匹配代价计算与聚合以及视差预测和细化四个阶段构成。然而，最近的研究工作都聚焦于前三个阶段，只有很少的工作将精力集中于视差细化之上[2] [3]。

受多任务学习在计算机视觉其它任务中日益成功的推动，立体匹配网络 SegStereoNet [4]以及EdgeStereoNet [5] [6]分别通过整合语义信息和边缘信息到网络之中，使得最终生成的视差图精度在整体或边缘结构区域上取得很大提升。在 StereoNet [7]中，使用图像信息来指导视差细化，并提出边缘感知上采样，Khamis 等人利用此方法成功从低分辨率初始视差图中生成高精度视差图，同时恢复出在下采样过程中损失的精细细节。而在 StereoDRNet [8]之中，Chabra 等人先使用图像信息和视差图信息分别计算出几何误差图和光度误差图，再用于指导视差细化，也取得了较大的精度提升。由此可以看出，在细化阶段引入额外的有意义信息指导视差细化是可进一步提高细化的结果，从而获得更高精度的视差图。

我们提出一种新的细化方法，其根本思想是在细化阶段把低分辨率视差图上每一点都作为一个视差平面，并在平面之上附加如局部代价卷等其它信息，它通过将已被证明对视差不连续处很有效的自适应卷积引入到视差细化中来，在相似视差平面上执行更新预测。在视差图上采样到更高分辨率的过程中，我们提出视差自适应采样，来实现倾斜视差平面传播，以避免由双线性插值上采样造成的精度损失问题。

总而言之，我们的贡献可总结如下：

利用自适应采样将聚合操作完全转换至视差域，等效实现大窗口的聚合操作而不会造成边界模糊。

提出自适应细化模块，使用自适应卷积在相似视差平面上执行更新预测操作。

提出视差自适应采样机制，基于倾斜视差平面上采样视差。

2. 相关工作

2.1. 视差注意力模块

我们的网络是基于 Wang 等人所提出的 PASMNet [9]架构得出的，我们只对视差细化模块进行替换，其它模块包括损失函数都没做任何改变。在该小节中，我们将简单介绍该网络并解释选择其作为基础网络的优势。若想要详细了解它，我们建议直接阅读原论文。

PASMNet 中最关键的模块为视差注意力模块，其基础原理是根据特征相似性沿极线计算左图像上一

点与右图像上一行所有点的注意力权重，从而捕获立体对应关系。如图 1 所示，将由特征提取网络生成的左右特征图 $A,B \in R^{H\times W\times C}$ 分别送入一个 1 × 1 卷积层中，以生成查询特征图 $Q \in R^{H\times W\times C}$ 和关键特征图 $K \in R^{H\times W\times C}$，并将 K 重整形为 $R^{H\times C\times W}$。随后，再对 Q 和 K 执行矩阵乘法运算，并应用 softmax 层计算出注意力权重，得到从右图像到左图像的视差注意力图 $M_{B\to A} \in R^{H\times W\times W}$。视差注意力模块还具有很强的通用性，简单调换 A、B 位置就能得到从左图像到右图像的视差注意力图 $M_{A\to B} \in R^{H\times W\times W}$。由于被遮挡区域的像素在右图像中找不到对应关系，它们在视差注意力图中往往被分配较小的权重，故只需对视差注意力图设置一个阈值 τ 并将不大于它的注意力权值都置为 0，就可获得一个有效掩码 V，如图 2 所示。

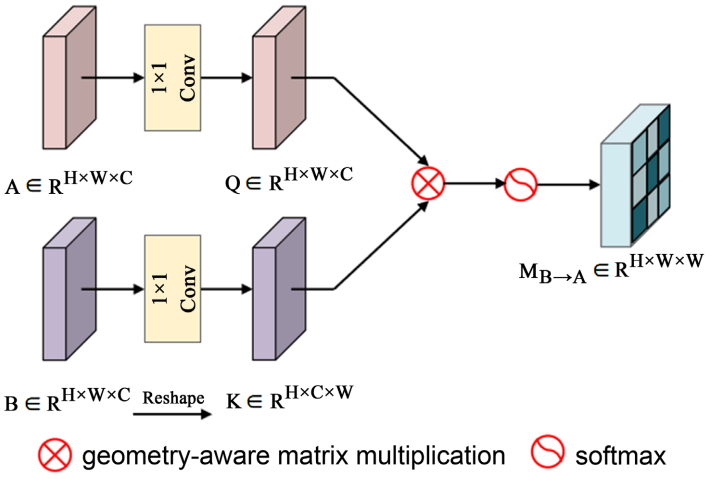

Figure 1. Illustration of PAM
图 1. PAM 示例

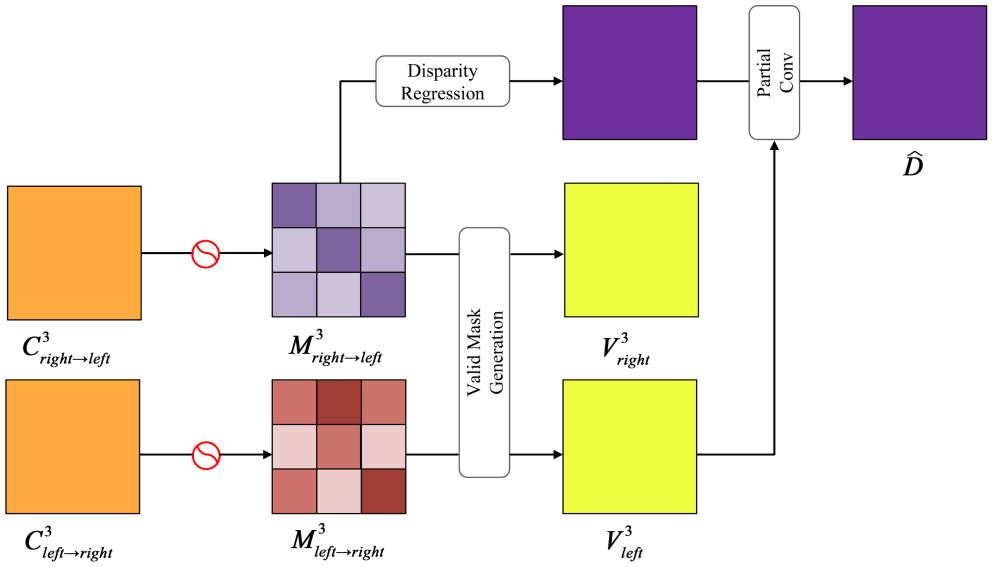

Figure 2. Output module
图 2. 输出模块

大多数基于 CNN 的立体匹配网络都使用代价卷技术来捕获立体对应关系[10] [11] [12] [13]。但是，具有不同基线、焦距和分辨率的立体相机在不同环境中获取的立体图像对中的最大视差可能会有很大差别，所以代价卷技术中使用的固定最大视差限制了这些网络在更复杂环境中的应用。然而，视差注意力

模块是基于极线约束和特征相似性通过矩阵乘法来获得不同的视差。因此，视差注意力模块拥有视差自适应的能力，可以处理较大的视差变换。

除以上优势之外，各个视差注意力模块之间还可以进行隐式聚合，如图 3 所示。同时，各个视差注意力模块之间通过残差连接紧密联系，这进一步提高了聚合效果。另外，对于低分辨率和高分辨率代价卷之间的融合，该网络除使用三线性插值上采样代价卷之外，还上采样了低分辨率特征图，并且使用一个 1 × 1 卷积层来融合高分辨率特征图和上采样结果。因此，低分辨率特征图也参与高分辨率匹配代价卷的生成与聚合，更有效地融合了不同分辨率的匹配代价卷。

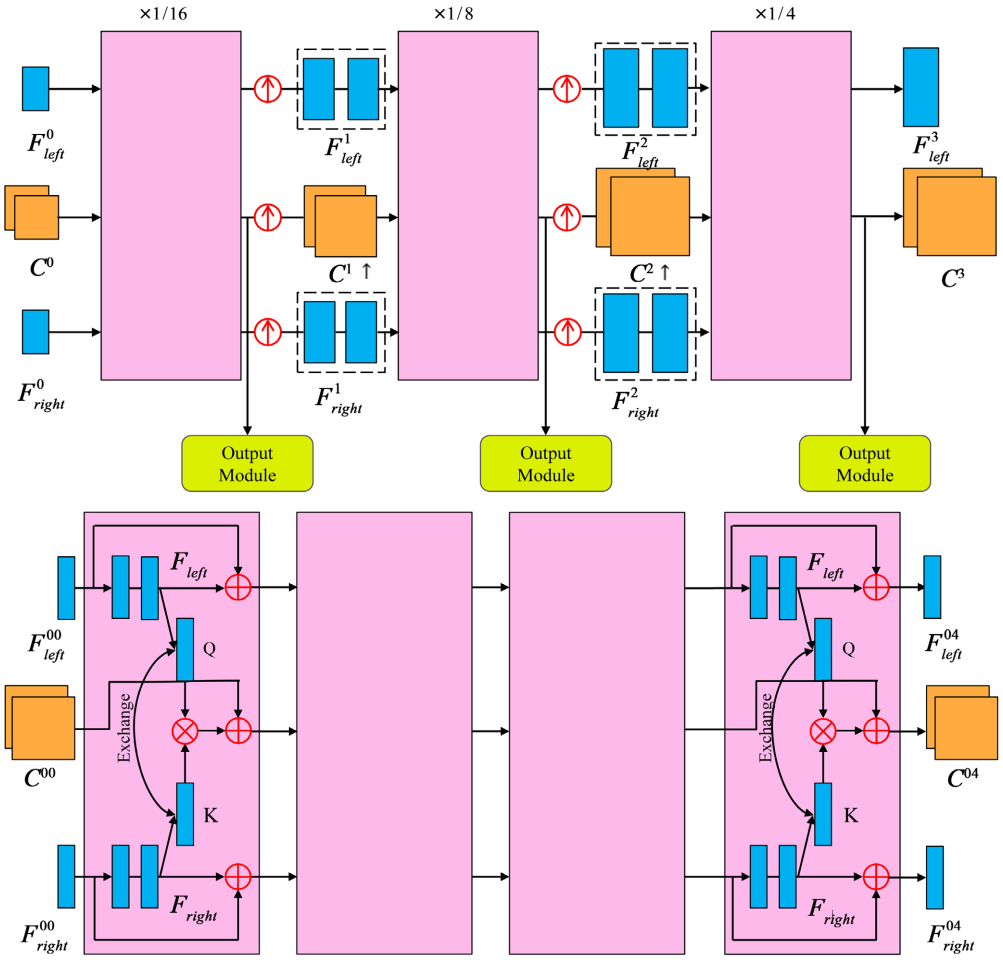

Figure 3. Parallax-attention module
图 3. 视差注意力模块

2.2. 自适应聚合

最先使用自适应聚合方法的工作是 Xu 等人提出的 AANet [14]，他们观察到局部聚合方法在视差不连续性处容易造成边界模糊，而且常规卷积存在权重与内容无关的缺点。为解决这些问题，他们将代价卷分别送入两个单独卷积层中，以得到用于自适应采样的额外二维偏移和用于内容自适应的权重。在 AANet 基础上，Wang 等人在 PatchmatchNet [15]中，分别从特征相似性和深度相似性两方面来计算自适应权重。在本文中，对于不同的任务，我们根据代价相似性、特征相似性、深度相似性中的一种或多种来计算自适应卷积的权值。

2.3. 贴片假设

Tankovich 等人在 HITNet [16]中使用 4 × 4 卷积层将特征图上 4 × 4 大小的图像平面转换为一个贴片，再利用快速高分辨率初始化步骤预测出贴片视差 d 以及用于附加额外信息的贴片特征描述符，这些附加信息完全由网络从数据中学习得出，没有做任何的约束或限制，它们可以描述这个平面的几何属性或预测视差的置信度或其它重要性质。然而，视差 d 只是平面中心处的视差，若想得到整个平面的视差还需 x 和 y 方向的视差梯度 dx、dy。因此，他们将以上所有信息都连接在一起并命名为贴片假设。在传播步骤中，基于信息传播和信息融合不断上采样并更新贴片假设直到分辨率大小与原始图像分辨率相同。其中，根据平面方程上采样视差 d，以实现倾斜平面传播，其余部分使用最近邻采样完成上采样操作。

3. 方法

3.1. 细化假设

受贴片假设启发，我们将对平面细化的思想与自适应卷积相结合，提出一种自适应的视差细化方法。由特征提取阶段生成的特征图上每一处都可以被看作是原始图像上一个 $2^s \times 2^s$ 大小的图像平面经 2^s 次下采样所得到的，所以每一点都表示一个平面。因此，在细化阶段，我们直接把每个点当作一个平面进行细化。如图 4(a)所示，除了预测视差 d，我们也为平面引入附加额外信息的特征描述向量 r，并将它们的连接结果定义为细化假设 h，

$$h = [d, r] \tag{1}$$

特征描述向量是通过卷积从左特征图和匹配代价中学习出的。如图 4(b)所示，我们利用输出模块预测的视差和特征提取模块得到的左右特征图按式(2)计算出每个像素 p 的匹配代价，

$$c(p) = \left\| F_L^{30}(p) - F_R^{30}(p-d) \right\| \tag{2}$$

获得匹配代价 $c(p)$ 之后，将其与左特征图相连后，送入一个 1 × 1 卷积层中，以生成特征描述向量。在上述步骤中，我们使用的左特征图是经视差注意力模块聚合处理后输出的，而不是直接使用特征提取阶段中生成的。因为与后者相比，前者融合了更多来自较低阶段的低分辨率信息，这使其在无纹理和弱纹理区域上更具判别性。

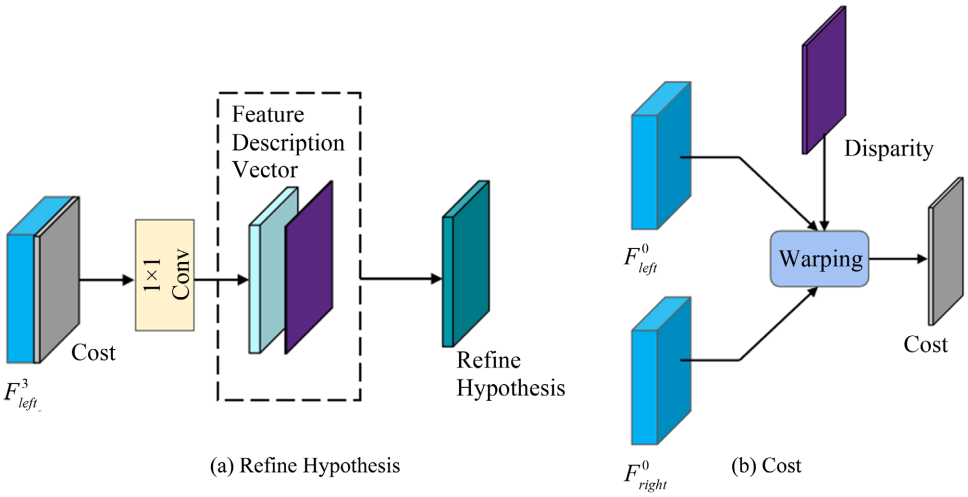

Figure 4. Overview of refine hypothesis
图 4. 细化假设概述

特征描述向量不仅允许将对视差细化有意义的额外信息附加到平面之上，还可以在自适应细化模块和视差自适应采样中为预测二维偏移和自适应权值提供来自特征图的信息。

3.2. 更新细化假设

在这一部分，我们先利用自适应传播网络构建一个局部代价卷来增强细化假设中所含的信息，而后再使用自适应细化模块来聚合增强细化假设，以在视差相似或相同的视差平面上预测出细化假设的更新。

3.2.1. 局部代价卷

此部分的关键一步是如何从匹配代价中构建出局部代价卷。经过多次的观察并联合实际情况，我们发现特征信息相似的像素往往也具有相似的视差。然而，由于输出模块预测的初始视差 d 中可能存在较大误差，特征信息相似的像素也可能具有一组很不相同的视差，我们希望通过局部代价卷能够将那些预测准确或误差较小的像素视差传播到其它像素中。若网络能够准确定位出一组视差平面相似的像素，则像素 p 在这一组像素所具有视差下的匹配代价可直接使用相应处的匹配代价近似(视差相似的像素，特征信息往往也相似)。采样这种近似方法可以加快网络的推理速度，同时还能使用较大窗口的卷积层对局部代价卷执行聚合操作，而不会出现边界模糊问题。

如图 5 所示，我们使用 PatchmatchNet 中提出的视差自适应传播机制为每一点都预测一组视差值并使用上述近似方法得到它们的匹配代价。通过在左特征图上应用一个 3 × 3 卷积层，网络为每个像素 p 都学习一个二维偏移 Δp_o，以定位出具有相似视差平面的像素点。值得注意的是，出于使网络更鲁棒的目的，我们同样采用经视差注意力模块聚合后的左特征图作为输入。然后，我们将学习到的二维偏移 Δp_o 加到固定偏移 p_o 之上，利用可变形卷积网络[17] [18]从匹配代价中自适应采样出 16 个采样点，其具体过程如下式所示：

$$\varphi(p)=\left\{c(p),\cdots,c(p+p_o+\Delta p_o)\right\}_{o=1}^{16} \tag{3}$$

$\varphi(p)$ 是像素 p 的局部代价向量，它可以看作是该像素匹配代价扩展后的结果，在这基础之上，我们可以构建出一个局部代价卷。

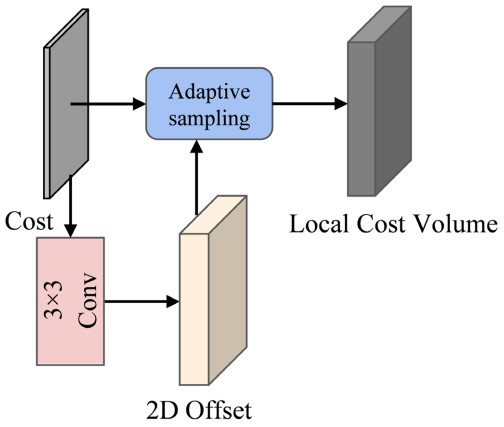

Figure 5. Local cost volume
图 5. 局部代价卷

出于使利用近似方法获得的局部代价卷更为合理的考虑，我们对局部代价卷执行代价聚合操作，此过程只需很少步骤和一个简单的 1 × 1 卷积层就可以实现。更详细地说，我们在网络中将初始视差 d 分别在正负方向上偏移一个视差单位恒定量得到 $d\pm1$，并且通过式(2)计算出这三个视差的匹配代价，以构建

出三个不同的局部代价卷。接下来，把它们按视差值大小的顺序连接起来，最后送入一个 1 × 1 卷积层中融合成单个局部代价卷，以完成代价聚合操作。通常来讲，不论使用 2D 卷积层，还是 3D 卷积层，代价聚合都是在空间域和视差域上同时进行的。然而，由于存在深度不连续的区域，使用窗口较大的卷积层来进行聚合会导致边界模糊问题。我们的方法将代价聚合操作只限制于视差域上，局部代价向量是由 16 个特征和视差平面都相似像素的匹配代价构成的，所以使用一个 1 × 1 卷积层来聚合局部代价卷就相当于使用一个窗口较大的 4 × 4 卷积层来进行聚合，但不会在深度不连续处出现模糊问题[19]，同时保留在无纹理或弱纹理区域更加鲁棒的优势。

3.2.2. 自适应细化模块

如图 6 所示，利用上一小节中聚合过的局部代价卷，我们对细化假设进行增强：

$$a(p)=\left[h,\varphi(p)\right] \tag{4}$$

其中，$a(p)$ 是增强细化假设，它包含了更多用于视差细化的信息。当预测视差有一定误差时，特征描述向量和局部代价卷都用于指导视差细化；当预测视差足够精确时，局部代价卷中含有的有意义信息较少，因为其中大部分元素都接近于 0，所以特征描述向量不受影响，依然可用于指导视差细化。

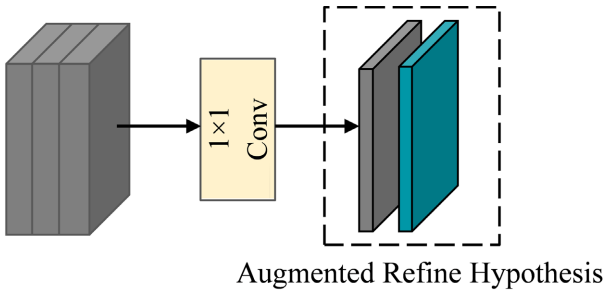

Figure 6. Augmented refine hypothesis
图 6. 增强细化假设

然后，我们使用参数为 $\theta_{\mathcal{U}_s}$ 的自适应细化模块 $\mathcal{U}_s$ 从相似的视差平面上对细化假设进行更新，如下：

$$\Delta r(p)=\mathcal{U}_s\left(a(p);\theta_{\mathcal{U}_s}\right) \tag{5}$$

我们提出的自适应细化模块只有一个 3 × 3 卷积层，它从增强细化假设中为每一个像素 p 都生成一个二维偏移量来定位出那些与之相似的像素，从而执行自适应聚合操作。与上一小节中只在左图像特征图上预测的 Δp_o 不同，从包含丰富信息的增强细化假设中得到的 Δp_u 精确性进一步提高。

根据生成的二维偏移 Δp_u 和固定偏移 p_u，我们对增强细化假设的特征通维中每一特征通道分别进行自适应采样，再把采样结果堆叠在一个新的维度上来让增强细化假设从三维张量转换至四维张量。接下来，我们将这个四维张量送入由卷积核为 1 × 1 × 1 的 3D 卷积层和 sigmoid 非线性层组成的权重预测网络中来预测出执行自适应聚合所需的权重。我们基于自适应卷积实现增强细化假设的自适应聚合，如下所述：

$$a(p)=\sum_{u=1}^{9} w_u \cdot a\left(p+p_u+\Delta p_u\right) \tag{6}$$

由于增强细化假设中包含预测视差、特征描述向量和局部代价卷信息，故 w_u 是权重预测网络根据深度相似性、特征相似性，代价相似性三个方面信息预测得出的。然而，与使用常规卷积进行聚合相比，我们的自适应卷积只在单一特征通道上执行聚合操作，在整个自适应聚合过程中都没有使用增强细化假设的其它特征通道信息，这导致用于自适应聚合的信息过少，对最后的输出结果有很大影响。因此，为

了预测出更精确的更新量 $\Delta r(p)$，我们在自适应卷积前后都使用一个 1 × 1 卷积层来融合特征通道信息，使自适应卷积拥有与常规卷积操作类似的效果，如图 7(a)中所示。最后，将经自适应聚合后的增强细化假设送入一个 1 × 1 卷积层中，以预测出 $\Delta r(p)$，将其与细化假设相加，完成一次更新操作。

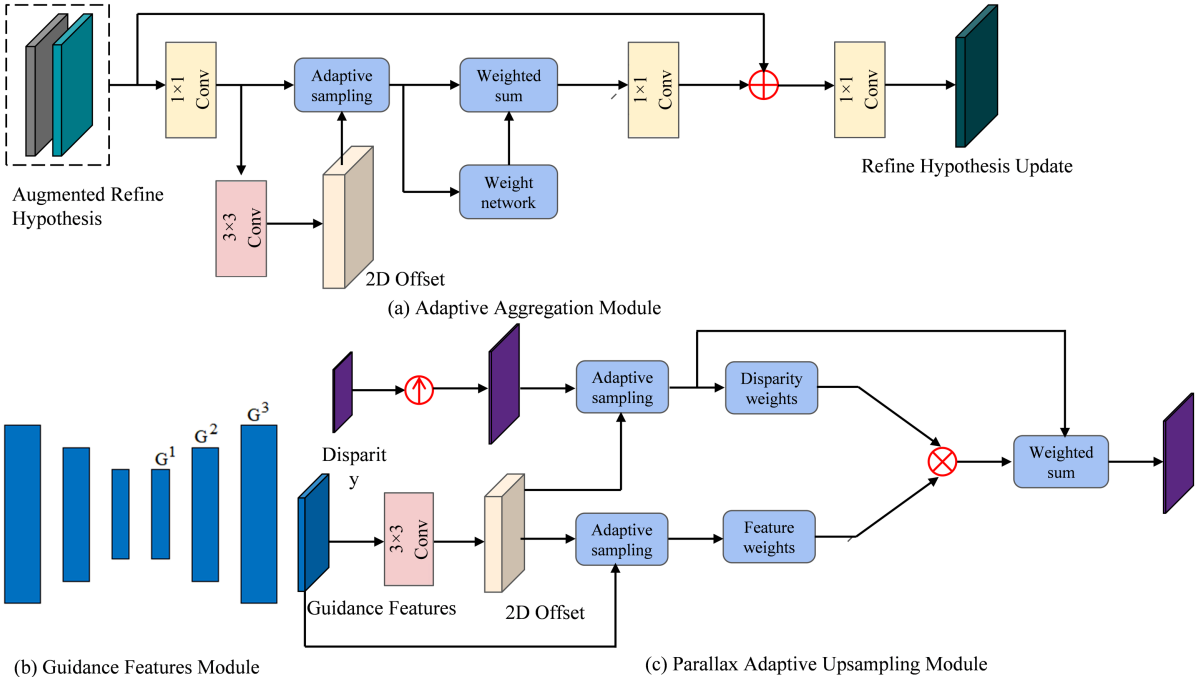

Figure 7. Overview of adaptive parallax refinement method
图 7. 自适应视差细化方法概述

3.3. 视差自适应采样

为了最终输出高精度视差图，我们没有将更新的细化假设直接上采样至全分辨率，而是使用一个分层细化网络逐层向全分辨率上采样并不断更新。出于高精度的要求，我们提出视差自适应采样机制上采样视差，即先利用双线性插值得到较高分辨率的视差图，再对其执行自适应聚合操作。

更新上采样的细化假设和使用视差自适应采样机制上采样视差都需要使用引导特征图。因此，我们修改 HITNet 中的特征提取器以用作引导网络，其架构类似于轻量网络 U-Net [20] [21]，即它含带有跳跃连接的编码器 - 解码器结构，如图 7(b)中所示。这表明，由引导网络生成的高分辨率引导特征图也包含一定程度的空间上下文信息，而且整个网络的计算量不大。我们将左图像送入引导网络中生成一组多分辨率的引导特征图 $\left\{g_L^0, g_L^{1/2}, g_L^{1/4}\right\}$。

与之前自适应方法类似，我们使用一个 3 × 3 卷积层从引导特征图中预测出一个额外的二维偏移 Δp_k，并且使用它与固定偏移 p_k 相加的结果来自适应采样出一组采样点，以计算出自适应聚合权重。然后，再按式(7)进行自适应聚合操作：

$$d(p) = \sum_{k=1}^{9} d_k g_k \cdot d\left(p + p_k + \Delta p_k\right) \tag{7}$$

其中，d_k 是深度权重，由视差权重网络根据深度相似性从使用双线性插值方法获得的上采样视差图中计算出；g_k 是引导特征权重，由引导特征权重网络根据特征相似性从引导特征图中计算出，如图 7(c)中所示。具体来说，视差权重网络较为简单只有一个 sigmoid 非线性层。而引导特征权重网络结构较为复杂，由三个 1 × 1 × 1 卷积层和 sigmoid 非线性层构成。

在增加自适应卷积之后，我们的网络在一定程度上拥有根据引导特征图调整视差大小的能力，这可以被看作为对上采样的视差图单独执行了一次细化操作。而对于特征描述向量，使用最近邻采样方法进行上采样即可。

跟随[16]，在分层细化网络中，我们只用引导特征图来增强上采样的细化假设并使用一系列具有不同扩张因子的残差块[22]来进行更新操作。此外，通过不同的上采样操作，细化假设中所包含不同的信息可以从低分辨率阶段向高分辨率阶段流动，保证网络具有灵活性。需强调的是，虽然视差注意力模块预测的视差图经自适应细化模块处理后已具有较高的精度，但更新过的细化假设在上采样之前，还使用上述方法再次进行更新，这一点尤为重要。不断重复这个过程直至每个点表示的平面大小为 1×1，我们会得到已更新的全分辨细化假设，这是网络的最终预测，从中我们可以输出全分辨率大小视差图。

4. 实验

4.1. 训练

为验证本文所提出的自适应视差细化方法的性能，本节将在 SceneFlow [12]和 KITTI2015 [23]两个数据集之上设计相关的对比实验。其中，SceneFlow 是合成数据集，训练集包含 168,357 个立体图像对，测试集包含 19,854 个立体图像对用于在训练阶段测试模型性能；KITTI2015 是从现实世界采集的数据集，最大视差为 192，训练集包含 400 张图像，验证集包含 800 张无真实标签的图像。对于所有数据集，为了保证训练时的显存爆炸问题，将输入图像裁剪为 512 × 256 的大小，并将图像的 RGB 值进行[−1, 1]之间的归一化。本节的实验所用到的软硬件实验平台如表 1 所示。

Table 1. The software and hardware platform required for high-precision algorithm operation
表 1. 高精度视差计算算法运行所需软硬件平台

平台内容	具体参数
CPU	Intel Xeon Silver 4116 @ 2.10 GHz
GPUNVIDIA TITAN	Xp 显卡
内存	DDR4 3200 Hz 128 GB
操作系统	Ubuntu 16.04 LTS
深度学习框架	PyTorch 1.7.1

在训练超参数设置部分，批处理大小设置为 4，最大视差值设置为 192。神经网络的权重优化使用 Adam 优化器，所有网络都首先在 SceneFlow 数据集上进行预训练，然后再 KITTI2015 数据集上进行迁移学习。

4.2. 实验对比

本文算法是基于 PAMSNet 基准模型之上所做改进得到的，为了突出我们的改进效果，算法性能继续使用 PAMSNet 的度量指标来进行评估，即端点误差(EPE)和 3 像素错误率(当一个像素的视差误差大于 3 个像素或大于自身真实值的 5%时，就可以认定它是一个错误像素)。在 SceneFlow 和 KITTI2015 两个数据集上的比较结果如表 2 所示。

通过实现分析数据得出，在 SceneFlow 数据集上，我们改进后的算法的 EPE 和 3 像素错误率分别超过原始算法 0.21%和 4.9%，而在 KITTI2015 数据集上，则分别达到 9.7%和 12.5%，结果如图 8 所示。

Table 2. Comparison of the improved algorithm with the original algorithm
表 2. 改进的算法与原始算法比较

Model	SceneFlow		KITTI2015	
	EPE	>3px	EPE	>3px
PASMNet	6.227	0.203	1.389	0.080
PASMNet_AR	6.214	0.193	1.254	0.070

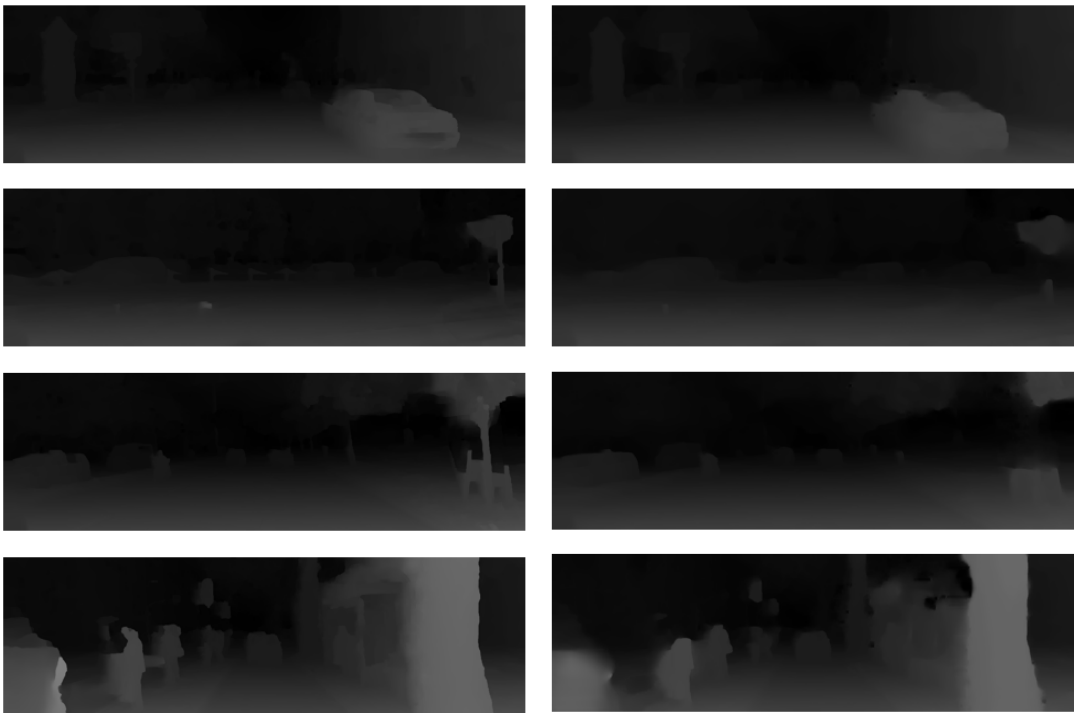

Figure 8. Experimental comparison results on KITTI2015, with the improved model on the left and the original model on the right
图 8. 在 KITTI2015 上实验对比结果，左边为改进模型，右边为原模型

5. 结论

本文提出了一种基于自适应卷积实现的视差细化与采样方法，将视差图上每一点都当作一个视差平面，在视差细化阶段为其引入额外对细化有意义的信息，利用自适应卷积在相似视差平面上更新视差，并使用视差自适应采样方法提高低分辨率视差图双线性插值上采样的精度。当在较低分辨率阶段时，分别使用局部代价卷和引导特征图进行信息增强；当在较高分辨阶段时，只使用引导特征图进行信息增强。在 SceneFlow 和 KITTI2015 数据集上，与原始算法相比较，两项指标都获得了提升，特别是在 KITTI2015 数据集上，效果尤为明显，达到 9.7%和 12.5%。

参考文献

[1] Scharstein, D. and Szeliski, R. (2002) A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, **47**, 7-42. https://doi.org/10.1023/A:1014573219977

[2] Poggi, M., Tosi, F., Batsos, K., *et al*. (2021) On the Synergies between Machine Learning and Binocular Stereo for Depth Estimation from Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 5314-5334. https://doi.org/10.1109/TPAMI.2021.3070917

[3] Laga, H., Jospin, L.V., Boussaid, F., *et al*. (2020) A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 1738-1764.

[4] Yang, G., Zhao, H., Shi, J., *et al*. (2018) SegStereo: Exploiting Semantic Information for Disparity Estimation. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 636-651. https://doi.org/10.1007/978-3-030-01234-2_39

[5] Song, X., Zhao, X., Hu, H., *et al*. (2018) EdgeStereo: A Context Integrated Residual Pyramid Network for Stereo Matching. In: *Asian Conference on Computer Vision*, Springer, Cham, 20-35. https://doi.org/10.1007/978-3-030-20873-8_2

[6] Song, X., Zhao, X., Fang, L., *et al*. (2020) EdgeStereo: An Effective Multi-Task Learning Network for Stereo Matching and Edge Detection. *International Journal of Computer Vision*, **128**, 910-930. https://doi.org/10.1007/s11263-019-01287-w

[7] Khamis, S., Fanello, S., Rhemann, C., *et al*. (2018) StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 573-590. https://doi.org/10.1007/978-3-030-01267-0_35

[8] Chabra, R., Straub, J., Sweeney, C., *et al*. (2019) StereoDRNet: Dilated Residual StereoNet. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 11786-11795. https://doi.org/10.1109/CVPR.2019.01206

[9] Wang, L., Guo, Y., Wang, Y., *et al*. (2020) Parallax Attention for Unsupervised Stereo Correspondence Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 2108-2125.

[10] Kendall, A., Martirosyan, H., Dasgupta, S., *et al*. (2017) End-to-End Learning of Geometry and Context for Deep Stereo Regression. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 66-75. https://doi.org/10.1109/ICCV.2017.17

[11] Chang, J.R. and Chen, Y.S. (2018) Pyramid Stereo Matching Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 5410-5418. https://doi.org/10.1109/CVPR.2018.00567

[12] Mayer, N., Ilg, E., Hausser, P., *et al*. (2016) A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 4040-4048. https://doi.org/10.1109/CVPR.2016.438

[13] Jie, Z., Wang, P., Ling, Y., *et al*. (2018) Left-Right Comparative Recurrent Model for Stereo Matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 3838-3846. https://doi.org/10.1109/CVPR.2018.00404

[14] Xu, H. and Zhang, J. (2020) AANet: Adaptive Aggregation Network for Efficient Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 14-19 June 2020, 1959-1968. https://doi.org/10.1109/CVPR42600.2020.00203

[15] Wang, F., Galliani, S., Vogel, C., *et al*. (2021) PatchmatchNet: Learned Multi-View Patchmatch Stereo. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 14194-14203. https://doi.org/10.1109/CVPR46437.2021.01397

[16] Tankovich, V., Hane, C., Zhang, Y., *et al*. (2021) HITNet: Hierarchical Iterative Tile Refinement Network for Real-Time Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 14362-14372. https://doi.org/10.1109/CVPR46437.2021.01413

[17] Dai, J., Qi, H., Xiong, Y., *et al*. (2017) Deformable Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 764-773. https://doi.org/10.1109/ICCV.2017.89

[18] Zhu, X., Hu, H., Lin, S., *et al*. (2019) Deformable ConvNets V2: More Deformable, Better Results. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 16-17 June 2019, 9308-9316. https://doi.org/10.1109/CVPR.2019.00953

[19] Min, D., Lu, J. and Do, M.N. (2011) A Revisit to Cost Aggregation in Stereo Matching: How Far Can We Reduce Its Computational Redundancy? 2011 *IEEE International Conference on Computer Vision*, Barcelona, 6-13 November 2011, 1567-1574. https://doi.org/10.1109/ICCV.2011.6126416

[20] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[21] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

[22] He, K., Zhang, X., Ren, S., *et al*. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[23] Menze, M. and Geiger, A. (2015) Object Scene Flow for Autonomous Vehicles. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3061-3070. https://doi.org/10.1109/CVPR.2015.7298925