

基于机器视觉的人脸关键点检测技术

刘俨 秦运柏 林中岚 何润卓 杨毅

(广西师范大学电子与信息工程学院 广西 桂林 541004)

摘要: 人脸关键点检测在众多人脸应用中扮演着重要角色,特别是在受限场景下的人脸识别、人脸验证和人脸矫正方面,这一技术一直是计算机视觉领域的研究热点。为了满足高效的人脸关键点检测需求,本文提出了一种基于改进YOLOv8的自由锚点人脸关键点检测算法。模型中加入了GAM全局注意力机制,沿用了不同模型尺寸的检测器,从实现最佳性能的大型模型到用于嵌入式或移动设备上实时检测的超小型模型。由于原来的L1, L2损失函数不利于训练后期的精度提升,使用Wing Loss损失函数,在训练的所有时期都能有效地检测回归。为了进一步提高准确度,本文提出了一种利用热力模型辅助坐标回归模型的方法。在热力辅助实验中,相比于未使用热力辅助的情况,准确度提高了3%,这进一步验证了算法的有效性。

关键词: 人脸检测; 机器视觉; YOLOv8; 注意力机制; Wing Loss 函数

0 引言

在计算机视觉领域,人脸识别技术是一个经典问题,引发了广泛的深入研究。该领域涵盖了人脸检测、人脸定位、人脸分割检测和人脸识别等一系列相互关联、密不可分的研究方向。这些研究方向共同构成了完整的人脸识别技术体系,为人们在图像和视频数据中准确、高效地识别和理解人脸提供了重要的技术支持^[1]。这些技术的应用已经极大地方便了人们的生活,尤其在轨道交通、通勤、身份验证等领域广泛应用。

在高人流密集的地下通道、车站出入口以及商场入口等场景中,高效实时的人脸检测和识别显得尤为重要。因此,人脸识别技术的研究具有非常高的经济价值和实用价值,并在多个领域带来广泛的应用和益处。它为安全管理和人员流动监测提供了强有力的支持,显著提升了公共场所的管理和服务水平。随着科技的不断发展,人脸识别技术将持续蓬勃进步,为社会带来更多便利和安全。这些前景令人振奋,并且对这一领域的持续探索和创新充满期待^[2]。

人脸关键点检测在人脸研究任务中十分重要。关键点能准确标注人脸图像中的眼睛、鼻子、嘴巴等部位,揭示了人脸的关键信息。这对于提升人脸技术性能,推动人脸研究的发展至关重要。随着技术不断进步,期待人脸关键点检测能在人工智能和人

脸应用领域持续发挥作用,为社会带来更多便利和智能化进步。

为了针对下文中的关键点检测问题进行研究,提出了一种基于改进的YOLOv8自由锚点人脸关键点检测算法。该算法通过检测头输出人脸标记框检测、关键点位置坐标以及人脸是否有遮挡等数据,从而实现对人脸关键点的准确检测和定位^[3,4]。通过结合YOLOv8简洁网络结构,引入了人脸关键点检测的全局关注机制,并对损失函数进行了修改,实现了快速的人脸检测和人脸关键点的实时检测和定位。下文以一种新算法为例进行说明。该算法在检测率方面具有显著优势,而且可以修改模型大小,以便部署在需要实时检测的场景和移动设备中。

1 相关工作

在计算机视觉初期阶段,人脸检测主要集中在受约束且特征明显的标准人脸图像上。而人脸关键点检测任务是通过处理完整的人脸图像,并找到包含特定语义的关键点坐标。根据现实应用场景的需求,通常根据关键点的不同数量划分模型,如17点、21点、68点和98点模型^[5]。无论目标关键点具体的数量,都要能够表征人脸的主要特征,如眼睛、鼻子、嘴巴和耳朵。现有的人脸关键点回归方法需要先在图片和关键点标注的数据集上进行训练,然后学习一个泛化能力强的计算方法,以便在未知图片上定

位关键点。目前的研究主要致力于提高关键点回归的精度和目标检测的准确率。

研究者们对人脸关键点检测提出了众多有效的方法,主要有传统方法和基于深度学习两个方法^[6]。传统方法包括基于主动模型和基于局部约束模型的经典技术。然而,在面对非约束的复杂环境、模糊、遮挡、倾斜、表情等情况时,传统的约束性人脸检测技术显得不够适应,难以胜任。基于深度学习的方法则利用卷积神经网络技术,从人脸图片中提取特征,进而准确定位人脸关键点的位置。这些方法涉及对数据集、模型和损失函数的各种优化方案,为人脸关键点检测的精度和效率带来了显著提升。通过深度学习技术的不断发展,可以期待在更加复杂多变的场景下,实现更准确、稳健的人脸关键点检测,为人脸研究和相关应用开辟更广阔的前景。

传统方法大致由基于模型的传统方法和基于局部约束模型的方法构成,Cootes等提出ASM^[1](Active Shape Model)和AMM^[2](Active Appearance Models)基于模型的传统方法^[7,8],作为早期的人脸关键点检测算法,它在图像处理中表现出很好的效果,并能够满足人脸检测的需求。基于局部约束模型的方法,是一种在人脸关键点检测中极具价值的技术。该方法综合考虑了全局的人脸形状信息,从整体上提高了检测的准确性和稳定性。同时,这一算法还突出了对每一个面部关键点周围局部特征的建模,使得关键点的定位更加精确和细致。优势在于综合了全局形状信息和局部特征建模,使得人脸关键点检测更具全面性和准确性。随着技术的不断进步,相信这一方法在未来会取得更加显著的突破和进步。

基于深度学习的方法在人脸关键点检测中按照回归的目标可以分成两类:热力回归和坐标回归。坐标回归法直接从图片回归出人脸关键点的坐标值。一种级联的三级CNN方法由Sun等人提出,它通过由粗到细的关键点检测过程,每个级别都由多个卷积网络组成,最终的全连接层输出坐标值^[9]。虽然取得了较好的性能,但它的复杂性过高,计算代价也较大。

另一个人脸监测算法为Cascade CNN^[2],其利用级联的多个检测网络来进一步优化估计结果。和一般的级联方式不同,Cascade CNN的几个检测网络都是根据对不同IOU阈值所确定的正负样本进行训练设计的。这个算法共使用了6个CNN的级联结构,进行了人脸锚点回归和边框调整。最后,通过应用

非极大值抑制方法来消除高度重叠的窗口,以得到最终的人脸检测结果。

尽管这两种方法都取得了一定的成功,但是第一种方法的复杂性和计算代价较高,对硬件资源要求较大,而第二种方法则在级联过程中取得了更好的检测结果。随着深度学习技术的不断发展,人脸关键点检测方法将会继续优化,并在更多应用场景中得到应用,为人脸相关的技术和应用带来更大的进步和发展。

在人脸关键点检测中,热力回归法是一种特殊的方法,它针对每个关键点分别生成对应的热力,并将这些生成的热力作为卷积网络回归的目标。相比之下,CALE提出了一种不同的两阶段卷积聚合模型,将得分图和早期的CNN特征结合在一起,从而得到回归热力的结果^[10]。另一方面,Yang等人采用了两个网络:一个用于校正面部图像,另一个堆叠了沙漏网络,用于生成预测的热力。而在处理多视角的人脸对齐问题上,JMFA则充分利用堆叠沙漏网络,取得了最佳的性能。此外,LAB方法通过显示估计,并结合面部边缘热力辅助回归关键点热力^[9]。尽管热力回归法的精度通常比坐标回归法更高,但它也存在一些不足之处,主要体现在需要更多的计算量和较大的模型大小,这使得与坐标回归法相比,其计算代价较高。因此,在实际应用中,需要综合考虑精度和计算资源之间的权衡,选择最合适的方法来满足具体需求。随着技术的不断进步,这些方法都将不断优化和发展,为人脸关键点检测领域带来更多的突破和进步。

在面对前述问题时,呈现了一种崭新的、基于YOLOv8模型的人脸关键点检测改进方法。这一方法通过独特的手法,利用预测框左上角顶点的偏移量来精准计算每个人脸关键点的位置,并将这些宝贵信息完美融合到最终的输出结果之中。为了确保模型成为独具匠心之作,对训练数据集进行了精心加工,在其中增添了详实而丰富的人脸关键点位置信息。这个过程不仅凸显了对细节的关注,也为模型的学习能力提供了坚实的基石。针对新增的关键点,引入了全局注意力机制,以增强对关键点的重点关注。

另外,为改善关键点的检测精度与可靠性,还将人脸关键点的Wing Loss损失函数导入了YOLOv8网络的训练中。通过对所有含有人脸关键点的置信

区域进行分析, 有效地达到了在输出人脸边界框的同时提供人脸关键点的效果。这一创新技术为关键点监测应用提供了更高的准确度和安全性, 也为人脸关联应用开辟了崭新的可能性。这一改进技术也大幅增强了人脸关键点监测的有效性, 这一创新方法为关键点检测领域带来了更高的精度和稳定性, 为人脸相关应用开启了全新的可能性。这一改进方法大大提升了人脸关键点检测的效果, 使得算法在复杂场景下具有更强的适应性和精准度。本文的研究对于推动人脸关键点检测技术在实际应用中的发展具有积极的意义。这一改进方法不仅使得检测到的人脸边界框更准确, 还使得关键点定位更加精确。

2 改进 YOLOv8 网络结构

2.1 主干网络

主干网络沿用 YOLOv5 经典架构的 Backbone+PANet+Head 的结构, 且 PANet 部分都是先上采样融合再下采样融合, 主干网络如图 1 所示。

Backbone 和 Neck 中采用的 C2F 结构, 其参考了 YOLOv7 的 ELAN 的设计思想, 用于替换 YOLOv5 中的 CSP 结构, 由于 C2F 结构有着更多的残差连接, 所以其有着更丰富的梯度流。

Head 部分不同, YOLOv5 是整体上输出的, 以 80 类为例, 因为每个像素点为 3 个锚点, 故每个像素点的尺寸为: $3 \times (4+1+80) = 255$; 而 YOLOv8Head 部分进行的结构操作, Cls 和 Box 是分开预测的, 并且从 Anchor-Based 换成了 Anchor-Free。

Loss 计算方面采用了 Task Aligned Assigner 正样本匹配策略, 并引入了 Distribution Focal Loss^[9]。

2.2 C2F 结构

C2F 结构由 CSP 结构改进而来。在 Backbone 中为 CSP1_X, X 个组件为残差结构, 其中两个 CBL 输入与输出的 Shape 是相同的。

YOLOv8 中的 C2f 结构如图 2 所示, 可以发现与 CSP 相比, 其增加了更多的跳层连接和额外的 Split 操作。

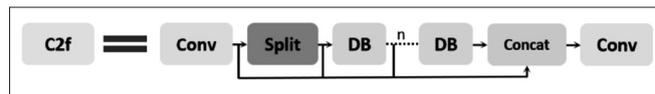


图 2 CSPayer_2Conv 结构

2.3 Head 从耦合头变为解耦头

从 YOLOv5 的耦合 Head, 变为解耦 Head, 并且由 Anchor-Based 变成了 Anchor-Free, 如图 3 所示。可以看出, YOLOv8 的 Head 中, 不再有之前的 Obj 分支, 只有解耦的分类和回归分支。同时, 从图 3 中可以看出, 解耦头中的 Cls 类别分支和 box 回归分支的通道数是不相等的, 因为各自表征了两种不同的特征。

(1) 对于 Cls 类别分支, 其通道数:

$$C_{cls} = \max(C_1, cls_{num}) \quad (1)$$

(2) box 回归分支其通道数:

$$C_{reg} = \max(16, C_1/4, 4 \cdot reg_{max}) \quad (2)$$

同时, YOLOv8 抛弃了 Anchor Box, 因为聚类 Anchor Box 是依赖于数据集的, 数据集如果不够充分, 就无法较为准确地反映数据本身的特征分布, 那聚类出来地 Anchor Box 就只能是次优的, 甚至更差。

2.4 损失函数

分类损失 Cls Loss, 使用 sigmoid 函数来计算每个类别的概率, 并计算全局的类别损失。训练标签中, 正样本的类别标签就是 IoU(Intersection over Union) 值; 负样本的标签全为 0。YOLOv8 采用了非对称的加权 VFL 损失函数, 公式如下:

$$VFL_{(p,q)} = \begin{cases} -q(q \log(p) + (1-q) \log(1-p)) & q > 0 \\ -ap^y \log(1-p) & q = 0 \end{cases} \quad (3)$$

式中: q - 标签, 正样本的时候 q 为预测 bbox 和 GT (Ground Truth) 的 IoU, 负样本时 $q=0$ 。

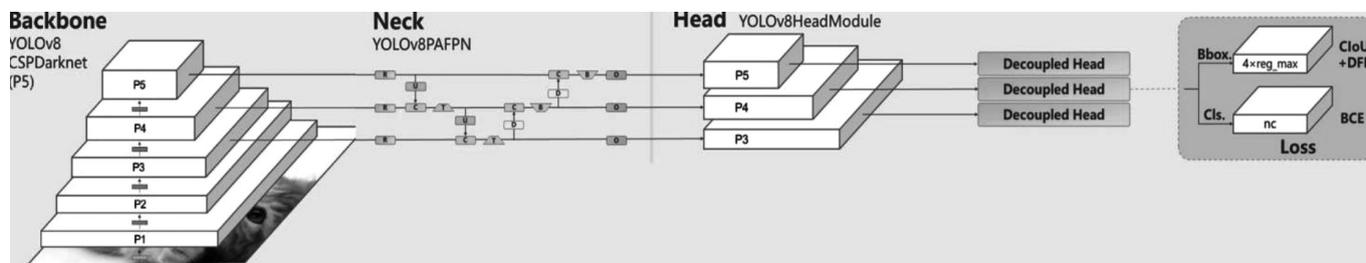


图 1 YOLOv8 网络结构

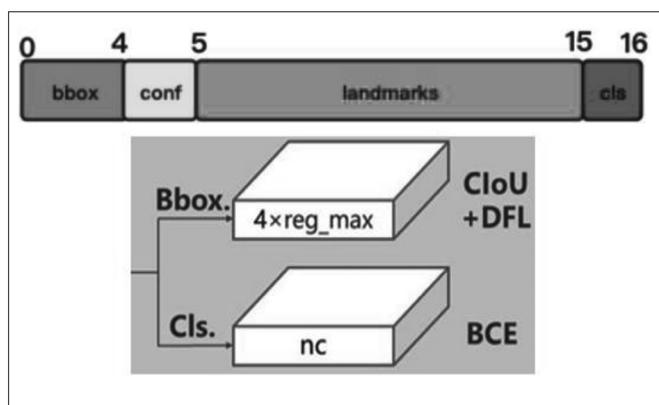


图3 Anchor-Based 变成 Anchor-Free

根据公式(3)可知,当为正样本的时候没有采用 Focal Loss,而是普通的 BCE Loss,只不过多了一个自适应的 IoU 加权,用于突出主样本。

3 算法改进

3.1 全局注意力机制 GAM

本文采用全局注意力机制 GAM (Global Attention Mechanism) 旨在有效地减少信息在全局维度上的弥散,并增强全局特征之间的交互。同时,还引入了序贯的通道-空间注意力机制,为模型赋予了更强的适应能力和表现力。然后对 CBAM (Convolutional Block Attention Module) 子模块进行了重新设计。在整个过程中,通道注意力图与空间注意力图的相互作用共同提取特征^[11],如图4所示。通过给定输入特征映射,定义了中间状态和输出。

$$F_2 = M_c(F_1) \otimes F_1 \quad (4)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (5)$$

这一综合优化的方法使得模型在处理复杂数据时表现出色,更加灵活高效。为了突出空间信息,在空间注意力子模块中采用了两个卷积层来融合空间信息。同时,在通道注意力子模块中,采用了与 BAM (Bottleneck Attention Module) 相同的缩减比例 r 。为了解决最大池化操作减少信息利用的问题,决定对注意力模块进行改进,删除了池化操作,以更好地保留特征映射。这一改进使得模型能够更充分地利用特征信息,提高了特征的表达能力和准确性。然而,需要明确的是,这样的调整可能会在一定程度上增加参数的数量,需要在设计中做出权衡。对模型的优化非常重要,寻求更优的设计方案,以平衡参数数量与性能之间的关系。尽管如此,这个

改进仍然在强调空间信息和提高特征提取性能方面发挥了积极的作用。

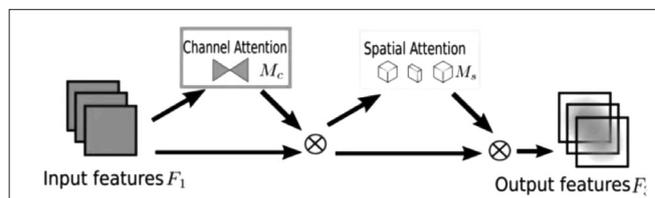


图4 全局注意力机制

3.2 Wing Loss 损失函数

在人脸检测和关键点回归任务中,检测和回归的任务难度不同,关键点之间的回归难度也不同。在初始化训练中,关键点回归损失很大,将其定义为 "large error";而在训练到中后期时,大部分关键点回归损失稳定在一个较小区域,将其定义为 "small error"。为了进一步提高检测回归的精度,必须增大损失函数的权重。在 Wing Loss 损失函数中,采用对数损失可以实现这一设想,其公式如下:

$$wing(x) = \begin{cases} w \ln(1 + |x|/\varepsilon) & \text{if } |x| < \omega \\ |x| - c & \text{otherwise} \end{cases} \quad (6)$$

这种类型的损失函数适用于处理由粗糙变精细的定位误差^[4]。

3.3 热力图辅助坐标回归模型

热力图是用来表示关键点在特征图中置信区域的图像。图像上某个位置的热力强度越低,代表着该点所在区域是置信域的可能性越小;相反,热力强度越高则表示该点所在区域是置信域的可能性越大。如图5所示,图中展示了对应于人脸关键点置信域的热力图,热力值最高的位置代表关键点置信度最大的位置^[3]。

热力图预测模型通常会为每个关键点生成一张热力图,然后通过计算每张热力图中亮度最大的像素位

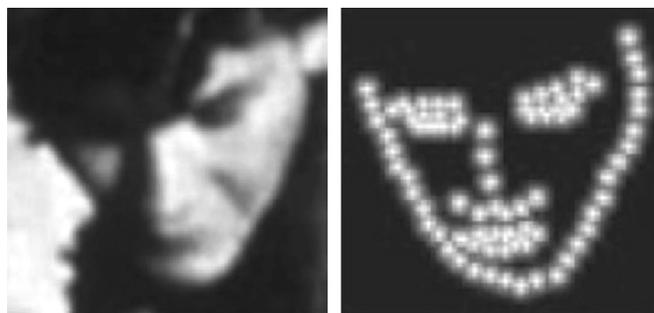


图5 关键点热力图

置来获得对应关键点的坐标。由于关键点的坐标不能直接得出,需要先生成热力图,再通过热力图计算坐标。因此,在梯度回流过程中,无法直接从坐标中传递梯度,而是通过间接的方式热力回流,这同时引入了更多的误差。另外,根据热力图计算回归坐标也会带来额外的计算开销和时间复杂度,导致性能下降。

4 实验分析

4.1 数据集

选用了主流数据集 WFLW (Wider Facial Landmarks in-the-Wild) 进行本研究。WFLW 数据集包含了 10000 张人脸图像,随机划分训练集和测试集,其中训练集包含 7000 张人脸,测试集包含 3000 张人脸。每张人脸图像都标注了 98 个特征关键点,除了关键点的标注外,该数据集还包含其他图片属性的标注,简化为遮挡、缺失和完整。相较于其他数据集,WFLW 数据集中涵盖了许多预测难度较高的图像,其中包括遮挡、表情和大姿态等特性。因此,该数据集具挑战性,对人脸关键点检测模型的鲁棒性和准确性提出了较高的要求。

4.2 测试指标

本文采用了三种性能指标来评估模型表现,分别是 NME (Normalized Mean Error)、Accuracy rate 和 AUC (Area Under the Curve):

(1) NME 是归一化平均误差,通过将 MSE (Mean Square Error) 除以归一化因子计算得出。一般情况下,归一化因子选取外眼角间距和瞳孔间距。

(2) 准确率 (Accuracy rate) 表示预测的准确性。若将 NME 大于 α 的测试样本视为预测失败,则准确率表示预测成功的样本占总测试集的比例。

(3) AUC 计算误差累积曲线下,误差从 0 到 α 的面积,然后再除以 α 。误差累积曲线上的每个点 (x, y) 代表着 NME 小于 x 的测试样本在测试集中所占比例为 y 。

在实现上,本文采用了特别适用于构建和训练各种深度学习模型 PyTorch 框架,并使用 Adam 优化器进行权重衰减(权重衰减率为 $1e-6$, L2 正则化)。每次调整学习率时,学习率将衰减为之前的 0.1,初始学习率设置为 $1e-3$ 。实验中的 GPU 环境为 NVIDIA GeForce RTX 3080。

4.3 实验结果

本文使用 WFLW 训练集进行训练,实验改进后

的模型效果。并且在最后综合所有模型进行对比,在测试集上检测出能达到的最优模型,实验结果如表 1 所示。

表 1 模 P 型投对员比缺表

Model	NME	AUC	Accuracy rate
YOLOv8	0.22	0.31	0.78
YOLOv8+GAM	0.17	0.26	0.83
YOLOv8+Thermal	0.15	0.21	0.85

实验结果表明添加全局注意力机制 (GAM),使用热力图辅助训练可以提升算法准确率。最优人脸关键点检测模型效果如图 6 所示。



图 6 检测结果

5 结语

为了克服当前人脸检测领域中锚框和人脸关键点回归能力弱的问题,本文充分发挥了 YOLOv8 模型在单阶段任务上速度方面的优势,并进行了相应改进。通过引入注意力机制、Wing Loss 损失函数以及热力图分析的设计,成功地实现了 YOLOv8 对人脸检测和关键点回归的同时预测。在保证人脸检测精度的同时,成功地实现了有效的人脸关键点回归。这些改进方法不仅提高了模型的检测性能,而且解决了人脸检测领域的关键问题,为人脸识别技术的发展贡献了重要的进展。

在本文中,针对现有的问题采用了一系列改进措施,充分发挥了 YOLOv8 模型的优势,以提高人脸检测和关键点检测的综合性能。通过这些改进方法不仅能够高效准确地检测人脸,还能输出关键点信息,为脸研究和应用领域带来了新的突破和可能性。

参考文献:

- [1] 朱望纯,张博.超轻量人脸关键点检测算法[J].电子测量技术,2023,46(05):98-104.
- [2] 王兴.基于循环推理网络的人脸关键点检测算法[J].现

代信息科技, 2023, 7(04):91-94.

[3] 付博闻, 李闯闯, 梁爱华. 基于改进YOLOv4-tiny的人脸关键点快速检测[J]. 计算机科学, 2022, 49(S2):450-454.

[4] 王小荣, 张益通, 彭炫. 基于人脸关键点检测的车载疲劳驾驶系统研究[J]. 无线互联科技, 2022, 19(17):82-84.

[5] 李丹阳. 基于深度学习的人脸关键点检测算法研究[D]. 武汉: 华中科技大学, 2021.

[6] 屈金山, 朱泽群, 万秋波. 基于YOLOv3的人脸关键点检测[J]. 长江信息通信, 2021, 34(01):55-58.

[7] Arnett G W, Bergman R T. Facial keys to orthodontic diagnosis and treatment planning. Part I[J]. American Journal of Orthodontics and Dentofacial Orthopedics, 1993, 103(04):299-312.

[8] Aboah A, Wang B, Bagci U, et al. Real-time multi-

class helmet violation detection using few-shot data sampling technique and yolov8[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:5349-5357.

[9] Terven J, Cordova-Esparza D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond[J]. arXiv preprint arXiv:2304.00501, 2023.

[10] 何卓骋. 人脸关键点实时检测的稳定化技术研究[D]. 合肥: 中国科学技术大学, 2020.

[11] 秦晓飞, 盛凯, 朱玥, 等. 基于注意力模型的人脸关键点检测算法[J]. 光学仪器, 2020, 42(02):45-49.

作者简介: 刘俨(2002.11-), 男, 汉族, 广东揭阳人, 本科, 研究方向: 自动化。

广告征订



版位 Format	价格 Price (RMB)
特殊版位 Specified Ads. Position	
封面	25,000
封二	16,000
封三	12,000
封底	18,000
扉一	15,000
扉二	10,000
后扉一	12,000
后扉二	9,000

版位 Format	价格 Price (RMB)
正常版位 Editorial Page	
编辑页	10,000
编辑页跨页	15,000
1/2编辑页	5,000
1/3编辑页	3,500
1/4编辑页	2,500

注: 所有特殊版位广告均为4C广告, 正常版位广告均为黑白色; 所有广告需提供成熟设计稿, 如需编辑部制作需单独收费。

优惠说明:

在原价格基础上, 连续预定3期, 优惠8%; 连续预定6期, 优惠15%; 连续预定12期, 优惠20%; 连续预定18期, 优惠30%; 连续预定36期, 优惠40%。另, 如提前一次性付款, 可在享受优惠的基础上享受8%的额外折扣。

广告预定热线: 010-6741 0664 / 1368 332 6370