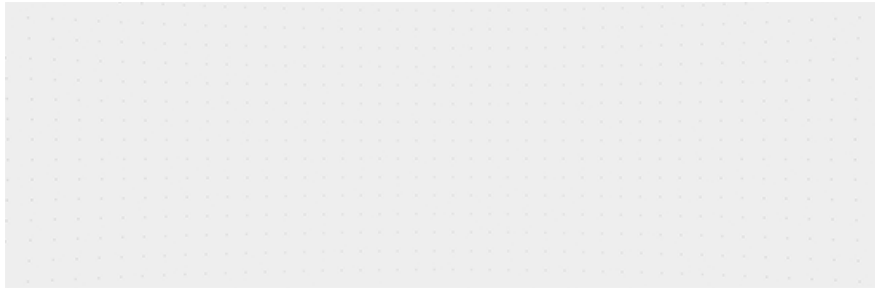


NTIRE 2021 冠军方案：华硕和台湾大学提出深度引导影像重照明方案-MBNet

原创 CV开发者都爱看的 极市平台 2021-05-30 22:00:00 手机阅读 罍

↑ 点击蓝字 关注极市平台



作者 | 杨皓翔、陈韦廷（华硕电脑、台湾大学）

编辑 | 极市平台

壹伴图

极市平台
extreme

月发文数目： **

月平均阅读： **

文章工具

已发文

采集图文 合成多

采集样式 查看

极市导读

华硕和台湾大学提出MBNet获NTIRE 2021 深度引导影像重照明冠军，该方案拿下了最高的SSIM和MPS并使用了track II的资料，让训练集数量翻倍。>>加入极市CV技术交流群，走在计算机视觉的最前沿

Multi-modal Bifurcated Network for Depth Guided Image Relighting

Hao-Hsiang Yang^{1*}, Wei-Ting Chen^{1,2*}, Hao-Lun Luo³, and Sy-Yen Kuo³

¹ ASUS Intelligent Cloud Services, Asustek Computer Inc, Taipei, Taiwan

² Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan

³ Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan.

(islike8399, jimmy3505090)@gmail.com, (r08921051, sykuo)@ntu.edu.tw

<https://github.com/weitingchen83/NTIRE2021-Depth-Guided-Image-Relighting-MBNet>

论文链接: <https://arxiv.org/abs/2105.00690>

NTIRE 比赛报告链接: <https://arxiv.org/abs/2104.13365>

代码链接: <https://github.com/weitingchen83/NTIRE2021-Depth-Guided-Image-Relighting-MBNet>

前言

最近在看CVPR2021中新 Trends in Image Restoration and Enhancement workshop的比赛，发现其中一个赛道，深度引导影像重照明(Depth Guided Image Relighting)很感兴趣，和典型的去噪，超分不太一样，而且还需考虑深度的空间信息，似乎是非常新的领域，于是看了一下比赛的报告和冠军方案的文章。冠军方案是由华硕和台湾大学联合提出的MBNet，网络设计和数据强化的方法挺有意思，遂纪录一些想法，欢迎交流。

简介

Depth-guided Image Relighting，字面上就是要输出不同光源的图片，不同光源可能指说不同方向光 and 不同色温。不同方向和不同颜色的光对于同样的场景，就会有不同方向的影子和明亮差异。举个例子，下面图(a)和图(d)分别是输入和输出，其中输入是向

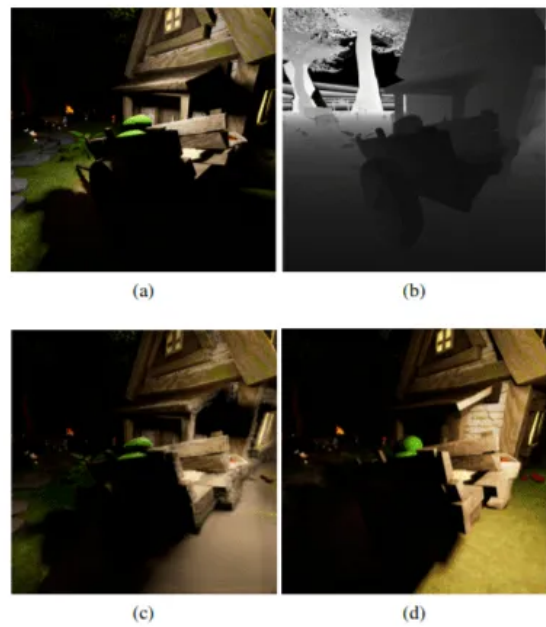


Figure 1: Example of depth guided image relighting. (a): Original input image. (b): Corresponding depth map. (c): Relighted image by our method. (d): Ground truth.

比赛规则

比赛使用了合成的数据集VIDIT (Virtual Image Dataset for Illumination Transfer)，共有390个场景，分成300个训练集，45个验证集，45个测试集。而对于每一个场景，提供5种色温和8个方向角的光照情况，所以一个场景，有40张不同光线情况的图和深度图。Track I的任务是输入图光线是色温6500K是朝北，输出为4500K是向东的光，而剩下的图片是用在Track II。最后计算模型在测试集的分，分数看结果图和GT的PSNR，SSIM，LPIPS，最后的指标是由LPIPS和SSIM综合的MPS。

冠军方案

网络架构

这里分成网络结构和数据强化的部分讲解。网络结构是多模态的叉型网络，为了要有效地融合图像信息和深度图信息，作者借鉴RGB-D SOD的思路设计模型，网络可以看成是典型的encoder-decoder结构，而图片通过网络学习特征后，作者使用残差学习，网络学习的是输出图和输入图的差。先介绍encoder端，在enocoder端，图像和深度图分别经过个别的backbone取不同尺度的特征，这里是选用2个ResNet50实现，而deocder端负责融合不同尺度的图像和深度特征。而在encoder端及decoder则是使用稠密联接(Dense connection)以及扩张动态金字塔模块(Dynamic dilated pyramid module, DDPM)来增加特征的传递能力，DDPM使用了Kernel generation units (KGUs)以及Kernel transformation units (KTUs)来增强特征的有效链接性，每个KGUs会使用融合的特征去产生不同的张量去涵盖不同的邻近感受野(3x3, 7x7, 11x11)，而KTUs则会产生规律的卷积模块搭配不同的扩张大小(Dilated rate)。decoder接收了DDPM的特征后，再逐步放大特征图，到最后层输出重照明图。

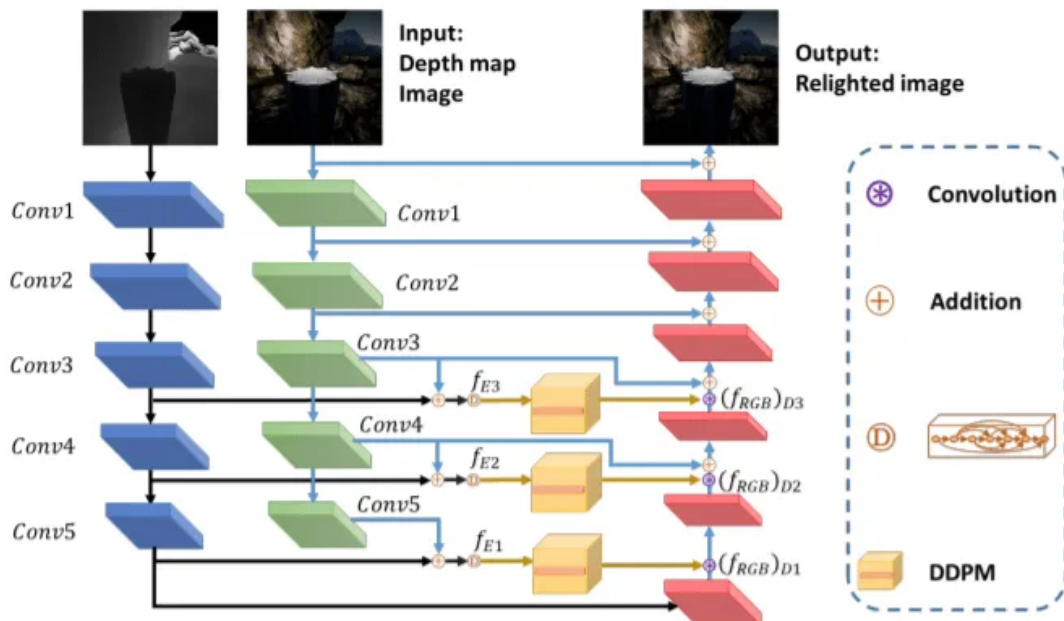


Figure 2: The architecture of the proposed multi-modal bifurcated network. The network consists of two streams as encoder parts: depth stream and RGB-image stream. We use the dense architecture ,DDPM and skip connection for better feature extraction as decoder parts.

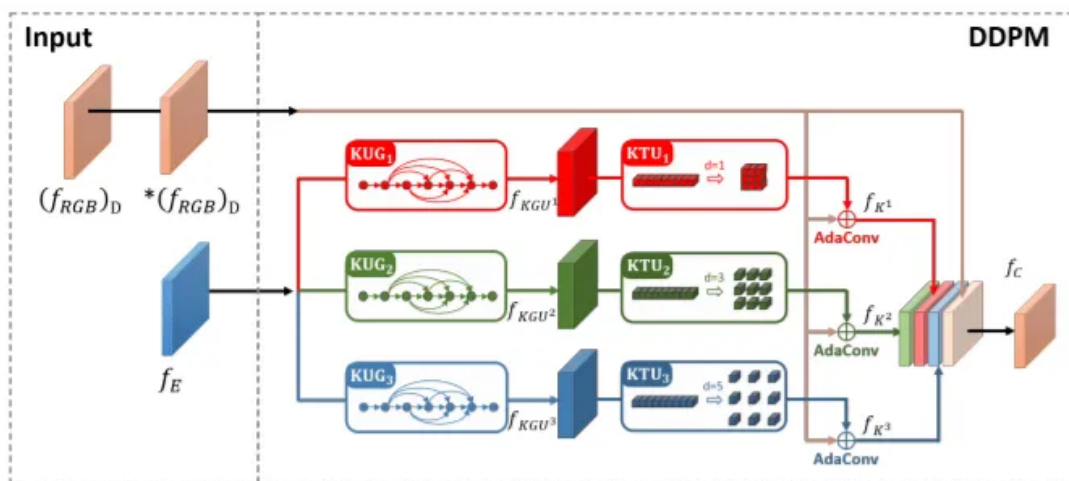


Figure 3: The architecture of the dynamic dilated pyramid module (DDPM). There are two modules in the network, that is, the kernel generation units (KGUs) and the kernel transformation units (KTUs).

网络训练的损失函数是采用Charbonnier loss、SSIM loss、perceptual loss三种损失函数，而图片本身就含有方向讯息，所以没有做随机翻转，裁切的操作，输入是一整张的图，要加强网络的鲁棒性，反而是用以下的方法进行数据增强。

数据强化

这里是我觉得论文最精明的地方，不只模型设计很重要，数据量也会决定模型的好坏。作者额外使用Track II的图片增加训练集数目，因为输入图的光线情况是色温6500K是向北的光，所以作者在实验阶段也使用色温6500K是向东北的光的图像，而输出图和深度图则维持一样，在下面的例图可以看的出来朝向东北和朝北的图片差别不大，这方法感觉简单粗暴，但硬是把数据量翻倍，在比赛中是很聪明的思路。

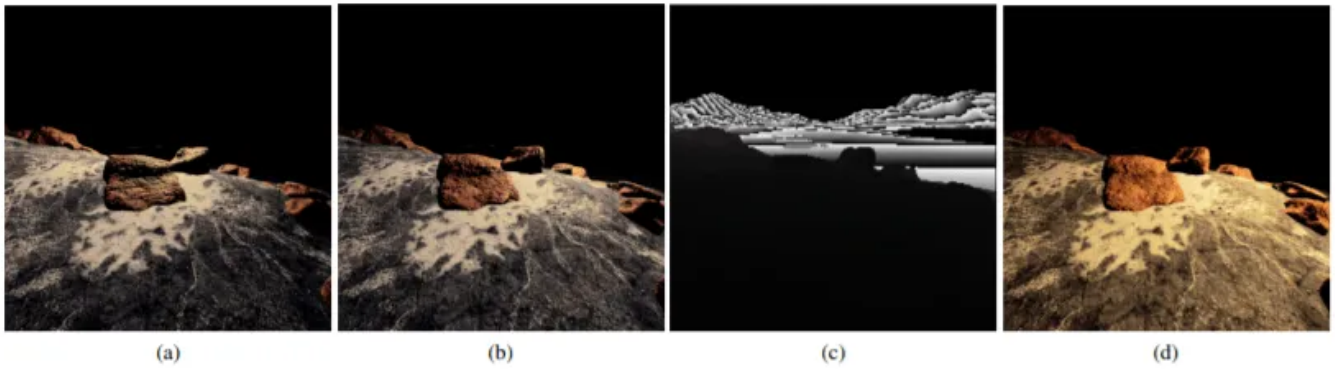


Figure 4: Additional images are used in the training phase. (a) Original image (6500-N). (b) Different illuminating angle image (6500-NE). (c) Guided depth map. (d) Output image.

实验结果

比赛主要计算结果图和GT的PSNR, SSIM, LPIPS, 最后的指标是由LPIPS和SSIM综合的MPS。作者先展示了使用额外track II的数据, 整体的分数有很大的改变, 其实很有道理, 毕竟训练集只有少少的300张, 模型很容易overfitting, 数据翻倍肯定让模型进步。再加上residual learning, 分数也又再进步。

Table 1: The ablation experiment of applying the different data and the residual learning.

Description	PSNR	SSIM
Baseline	18.0215	0.6834
+ Extra data	18.9677	0.7103
+ Extra data + Residual learning	19.3558	0.7175

接者又展示了比赛的结果。AICSNTU-MBNet拿下了最高的SSIM和MPS, 拿下了这个赛道的冠军。仔细看又发现LPIPS第一名的团队(NTUAICS-ADNet), 和冠军队伍是来自相同的单位, 真的牛逼。

Team	Author	MPS ↑	SSIM ↑	LPIPS ↓	PSNR ↑	Run-time	Platform	GPU
AICSNTU-MBNet	HaoqiangYang	0.7663	0.6931	0.1605	19.1469	2.88s	PyTorch	Tesla V100
iPAL-RelightNet	auy200	0.7620	0.6874	0.1634	18.8358	0.53s	PyTorch	Titan XP
NTUAICS-ADNet	aics	0.7601	0.6799	0.1597	18.8639	2.76s	PyTorch	Tesla V100
VUE	lifu	0.7600	0.6903	0.1702	19.8645	0.23s	PyTorch	P40
NTUAICS-VGG	jimmy3505090	0.7551	0.6772	0.1670	18.2766	2.12s	PyTorch	Tesla V100
DeepBlueAI	DeepBlueAI	0.7494	0.6879	0.1891	19.8784	0.17s	PyTorch	Tesla V100
usuitakumi	usuitakumi	0.7229	0.6260	0.1801	16.8249	0.04s	PyTorch	Tesla V100
MCG-NKU	NK_ZZL	0.7147	0.6191	0.1896	19.0856	0.33s	PyTorch	RTX TITAN
alphaRelighting	lchia	0.7101	0.6084	0.1882	15.8591	0.04s	PyTorch	Tesla K80
Wit-AI-lab	MDSWYZ	0.6966	0.6113	0.2181	17.5740	0.9s	PyTorch	RTX 2080Ti
Couger AI	Sabarinathan	0.6475	0.5469	0.2518	18.2938	0.015s	Tensorflow	GTX 1070

Table 1. NTIRE 2021 Depth-Guided Image Relighting Challenge Track 1 (One-to-one relighting) results. The MPS, used to determine the final ranking, is computed following Eq. (1).

最后展示重照明的结果, 展示前三名的结果, 看的出来模型的输出和GT差别还是挺大的, 可能是因为题目比较新颖, 困难, 整体的表现还有很大的成长空间。



Figure 1. A challenging example image from the NTIRE 2021 Image Relighting Challenge Track 1 (One-to-one relighting) with the output results of some top submission methods.

结语

读完这篇文章后，有两个特点非常值得学习

1，虽然是影像强化的工作，但作者也参考了RGB-D SOD的思想去设计网络，让我明白要多涉猎不同领域的论文，借鉴别的主题来加强原本的任务。

2，除了网络架构外，数据也是很重要的一环。现在NTIRE的其他赛道像是去雾，超分也都会用额外的数据集，而这篇论文用了track II的资料，让训练集数量翻倍，在比赛中是很聪明的操作。

参考文献

[1] M. El Helou, R. Zhou, S. Susstrunk, R. Timofte et al., "NTIRE 2021:Depth-guided image relighting challenge," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021.

[2] H.-H. Yang, W.-T. Chen, H.-L. Luo, and S.-Y. Kuo, "Multimodal bifurcated network for depth guided image relighting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021.

如果觉得有用，就请分享到朋友圈吧！



极市平台

专注计算机视觉前沿资讯和技术干货，官网：www.cvmart.net
624篇原创内容

公众号

▲点击卡片关注极市平台，获取最新CV干货
公众号后台回复“长尾”获取长尾特征学习资源～

极市干货

YOLO教程：一文读懂YOLO V5 与 YOLO V4 | 大盘点 | YOLO 系目标检测算法总览 | 全面解析YOLO V4网络结构

实操教程：PyTorch vs LibTorch：网络推理速度谁更快？ | 只用两行代码，我让Transformer推理加速了50倍 | PyTorch AutoGrad C++层实现

算法技巧（trick）：深度学习训练tricks总结（有实验支撑） | 深度强化学习调参Tricks合集 | 长尾识别中的Tricks汇总（AAAI2021）

最新CV竞赛：2021 高通人工智能应用创新大赛 | CVPR 2021 | Short-video Face Parsing Challenge | 3D人体目标检测与行为分析竞赛开赛，奖池7万+，数据集达16671张！

CV技术社群邀请函 #
.....

△长按添加极市小助手
添加极市小助手微信 (ID : cvmart2)

备注：姓名-学校/公司-研究方向-城市（如：小极-北大-目标检测-深圳）

即可申请加入极市目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解等技术交流群

每月大咖直播分享、真实项目需求对接、求职内推、算法竞赛、干货资讯汇总、与 10000+来自港科大、北大、清华、中科院、CMU、腾讯、百度等名校名企视觉开发者互动交流~

觉得有用麻烦给个在看啦~

阅读原文

喜欢此内容的人还喜欢

15个目标检测开源数据集汇总
极市平台