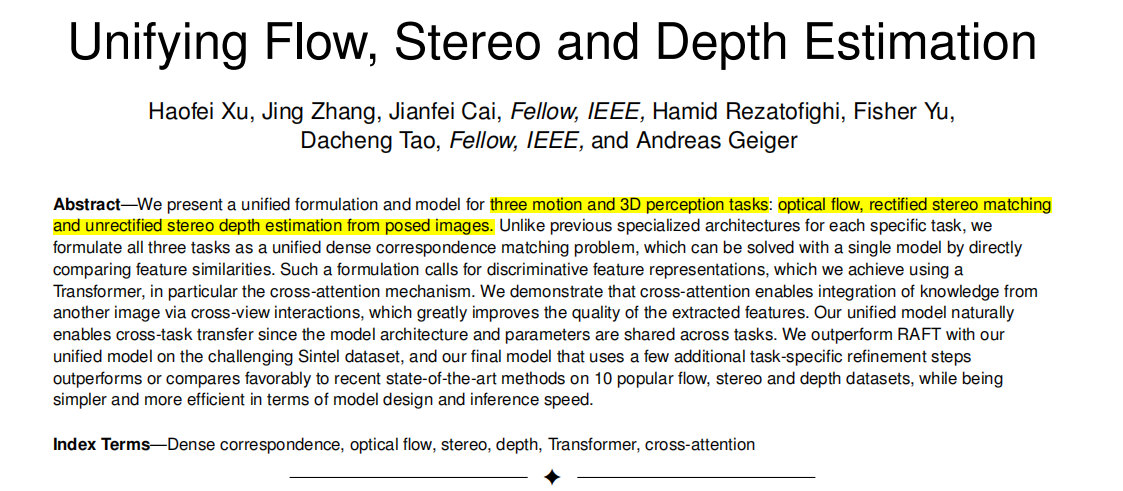
# 论文阅读报告：

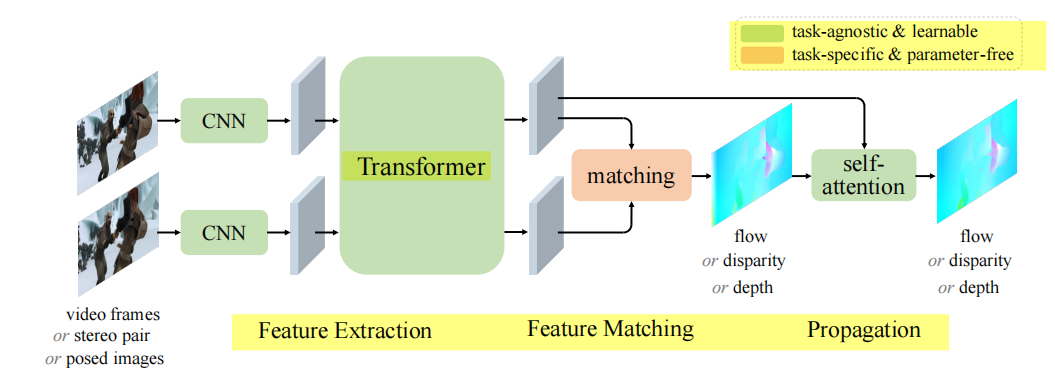
##### 一、



首先是这篇文章做了什么：

依照摘要，文章提出了一个基于transformer，能够解决optical flow、rectified stereo matching以及unrectified stereo depth estimation任务的统一模型方程。该模型具有跨任务转移能力（这是当前其它模型所不具有的统一性的体现），并在文章多处进行了对上述能力相应的意义论述以及优点的论文。

文章主要提出的统一模型采用如下架构进行构建：

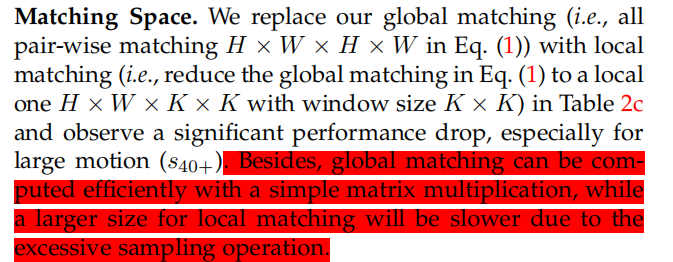


关于模型的介绍如下：

*模型考虑两个图像作为输入，可以是视频帧用于光流，立体图像对用于校正立体匹配，或摆姿势的图像用于非校正立体深度估计。首先使用权重共享的卷积网络从两个输入图像中提取8倍下采样的密集特征。然后，将这些特征输入到Transformer中进行特征增强（enhancement）。接下来使用无变参数的、基于特定任务的匹配层进行特征匹配，另外，引入了一个任务无关的自注意力层，通过测量特征自相似性将高质量的预测传播到未匹配的区域*

其中引入自注意层的目的在文章中也花了部分篇幅进行论述（考虑到存在单个输入相应区域存在被遮挡的情况，因此需要对引入self attention层通过领域信息的传播来解决）

文章分部分的介绍了上述三个部分的大致实现细节，其中花了较多的篇幅（2-3页）介绍了关于实现上述三个任务统一的方程推导过程，其中值得指出的是关于rectified stereo matching任务的推导，和之前我看过的大部分论文的实现思路不同的是这里采用了全局匹配思路（因为之前看过的主要的论文集中于双目视觉的局部匹配+卷积神经网络特征提取的策略），这里的全局匹配策略的使用的原因以及和局部匹配策略方法的优势在文中也有所论述：



即基于矩阵乘法的全局匹配方法实现能够有效提高大尺寸图片的处理速度。

同时在文章中探索了两种类型的细化：具有更高分辨率（1/4）特征的层次匹配细化和具有卷积的局部回归细化。层次匹配细化使用了基于匹配的公式（与任务无关）。而局部回归细化是特定于任务的，但可选的。因此，它可以看作是一个进一步提高我们的性能的统一方法的后处理步骤。

有意思的一点是这个统一模型在处理拥有两个输入较大位移差的时候的效果会比反之的效果更加明显，因此仍然需要局部CNN的特征提取进行后处理（post processing）。

综读整篇文章，统一性和与具体任务无关的可移植性是提出模型的相对于其它文章的优势，同时文章引入了交叉注意力机制用于更加充分的获得输入的上下文信息。同时，该模型的缺点包括：

*1. 在未匹配区域仍有改进空间。在Sintel数据集上，匹配区域的误差已经很小，但在未匹配区域误差较大，需要进一步研究。*

*2. 采用RAFT的迭代优化架构作为后处理步骤，以进一步提高性能。未来可以探索更轻量级、更有效的方法。*

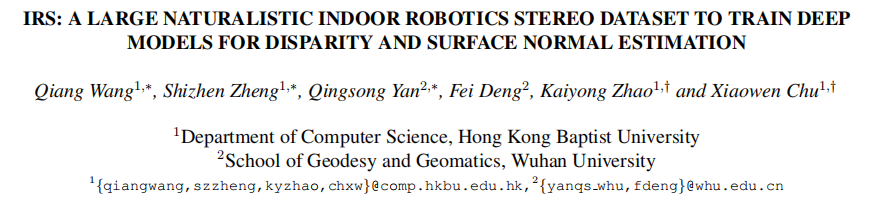
*3. 模型仍然无法实现实时推理速度。未来需要进一步改进，以适应具有实时要求的应用（20 FPS或更高）。*

个人感觉模型上述缺点并不是特别突出（毕竟是顶刊发布T-T，个人想法是模型的缺点仍然来自于全局匹配策略本身带来的部分不足（仅仅是不成熟的个人想法））

文章后面花了大篇幅介绍统一模型基于thing、setti测试集在解决上述三个任务的运行效果以及可移植性的优势分析（大概5页）。

##### 二、

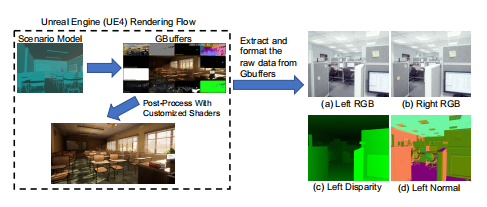
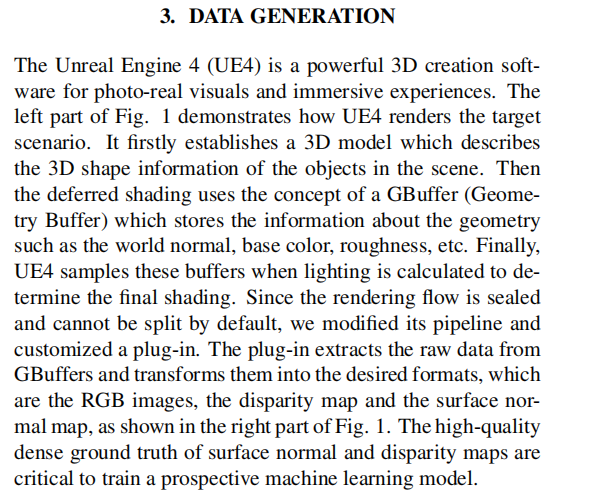
（将王强教授之前发给我的论文阅读了一篇）



文章提出了一个大型的自然室内机器人立体数据集（IRS），用于训练深度模型进行视差和表面法线估计。IRS数据集具有丰富的视觉属性，包括亮度变化、相机视野范围变化等（与真实世界场景非常接近）。本文将IRS与其他立体数据集进行定量分析并提出DTN-Net，用于预测室内场景的表面法线图。实验结果表明，IRS数据集在训练深度模型方面具有很好的效果，并且DTN-Net在预测表面法线图方面优于现有的方法。

文章主要介绍的是提出设计的数据集的来源以及对当前先进模型的训练效果，同时对于文章用于验证相关数据集提出的模型DTN-Net论述相关信息较少（个人猜测为了更加详细的介绍DTN网络的设计有另外一篇论文进行介绍，本文主要是介绍设计的室内机器人数据集IRS）。

阅读文章的目的在于了解数据集是如何生成的以及一个数据集合的好坏对模型训练会有什么样的影响，关于数据集的生成，论文中采用的是采用ue4引擎进行的场景搭建并使用延迟着色技术，通过几何缓冲区（GBuffer）存储有关几何的信息，如法线、基本颜色、粗糙度等。最后，UE 4在计算光照时采样这些缓冲区，以确定最终的着色。



这是我看过的第一篇关于模型数据集生成以及训练效果评估的论文.对于采用UE 4生成数据集这一种方法是没有想到的（对数据集产生是实景取景拍摄等方法获取），关于数据集的生成以及效果评判的论文阅读也有利于了解相关模型的无论训练集还是测试集的底层逻辑。