1. Problem Statement

本文试图解决的研究问题是什么?这个问题重要吗?为什么?

当前的技术状态(即其他竞争方法/解决方案)是什么?

这篇论文的动机是什么?这篇论文的定位是什么?

如果有必要，你可以用图片来更好地说明研究问题。

本文试图解决的问题是使用稀疏的输入，利用近场光捕获的图像高质量地恢复人脸的形状。这是一个计算机视觉和图形学中的一个重要问题，涉及到数字演员、人脸识别和动画等多种应用。

当前有一些方法使用多视图信息或照明条件（multi-view information or the illumination conditions）从源头解决该问题。其中有些方法可以给出高质量的结果，但是硬件环境难以建立，底层优化问题不易解决，计算成本较高。从单一图像进行人脸形状重建的方法大多侧重于 “in-the wild” image，但是这些方法很少能够恢复面部细节，并且大多由于阴影面积大不适合近点光照条件。基于深度学习的光度立体技术有两种方法，一种需要图像和校准光照条件，另一种直接从输入图像中估计光照条件并建模，但是这两种方法都侧重于特定方向照明条件下的建模问题且大多需要密集的输入，这通常难以实现。

针对这些问题，本文把基于深度学习的光度立体技术与人脸先验信息结合，在近场点光源照明下利用稀疏的输入得到高精度的三维人脸模型。

2. Summarise the paper’s main contributions

这篇论文声称有哪些贡献?

这篇论文有什么新内容吗?这篇论文有什么创新之处?

作者是否夸大了他们的贡献，如果是，又是如何夸大的?

作者们声称他们设计了一个轻量级的两阶段卷积神经网络，把基于深度学习的光度立体技术与人脸先验信息结合，可以在近场点光源照明条件下，用稀疏的输入进行高精度的三维人脸建模。该方法不需要严格的光源方向假设，而且仅需要1至3张近场点光源人脸图片作为输入即可工作，输入的图片越多结果越好。

同时由于缺少公开可用的数据集，他们构建了一个由三个近点光源和一个固定相机组成的系统，用于捕获真实的人脸图像作为数据集，并使用合成数据来增强数据集。

但是虽然他们说该方法是轻量级的，但是并没有给出与其他方法在处理时间上的对比。

3. Method and Experiment

他们提出了什么方法/想法/见解? 解释他们方法的关键(关键思想/洞察力/聪明才智)。

主要观点/论点是什么?

作者如何证实他们的主张?进行了哪些实验?这些实验的主要结果或发现是什么?

他们提出了一个两阶的卷积神经网络（CNN）。第一层网络以参数化三维人脸模型作为先验知识，利用位姿参数估计粗略的形状，并将获得的人脸图像和法线图输入第二阶段网络以生成更准确的法线图,最后通过一个fast surface-from-normal optimization恢复高质量的人脸模型。

他们使用组建的相机系统构建了一个由84名受试者提供的真实数据集，每个受试者有29种不同表情，并将从真实数据集中获得的反射率转移到随机生成的形状模型中。他们还基于Light Stage数据集构建了一个合成数据集，该数据集有准确的参数模型以供对照。

在参数化模型中，基于优化的方法将人脸近似为博朗曲面，在给定捕获的图像时，建模问题是恢复光源位置和照明强度，顶点位置，反射率和该点的法线。由于几何形状和反射率的模糊性，基于优化的方法非常耗时，且需要至少输入3张图像。因而他们提出了一种基于CNN的方法，从任意数量、不同近点光照条件的输入图像中学习。输入图像越多，效果越好。

在第一阶段，直接从单张图像中学习3DMM参数和pose参数，得到一个粗糙的模型。他们使用两个损失项，第一个computes the distance between the recovered geometry and the ground truth geometry：



第二个measures how close the projected 3D landmark vertices are to the corresponding landmarks in the imge:

钟表的特写

中度可信度描述已自动生成

最终的损失函数为：



其中Wlan是一个调优权重。

第二层网络结构类似于PS-FCN，由shared-weight feature extractor, an aggregation layer, and a normal regression module组成，但其normal估计网络使用proxy geometry作为输入，其损失函数为：

黑色的钟表

中度可信度描述已自动生成

where M is the set of all pixels in the face region covered by the coarse face model, ni and n^i is the estimated and ground truth normals at pixel i, respectively.

该方法与其他基于深度学习的光度立体视觉方法如UPSFCN和SDPS-Net相比较，该方法在定性和定量上都由于其他两种方法。该方法与其他单张图像人脸重建方法Extreme3D、DFDN相比较，细节恢复的更好，且平均几何误差更低，整体表现更好。

5. Critical Analysis

**5.1. Are the paper’s contributions** **significant?**

贡献/改进是微不足道的、增量的吗?

为什么之前的努力失败了?

论文的贡献是显著的。先前的基于优化的方法难以适应稀疏输入且计算困难。而基于CNN的方法缺少人脸参数化模型先验，需要特定方向的光照，而且在细节方面表现不佳。他们提出的将基于深度学习的光度立体视觉和参数化模型先验相结合的方法使得从图像恢复人脸形状更为方便可行，并且质量更好。

同时他们也扩充了数据集，为以后的相关工作提供了便利。

**5.2. Are the authors’ main claims valid?**

他们是否令人信服地证实了自己的主要观点?

他们的论点，推导，实验有漏洞吗?

作者的main claims是有效的。作者针对之前方法的局限提出了新的方法，给出了数学推导和粗略的网络模型，并通过实验对比了不同方法之间的差异，从最终的效果图片上看作者的方法确实有较高的性能。

**5.3. Limitation and weaknesses**

他们的方法有什么限制/弱点吗?可以做些什么来改进这项工作?

你会如何解决/克服他们的弱点?

从论文中的实验可以看出，当前的光源位置和拍摄角度较为固定，光源位置只有正面，左面和右面，并从正前方拍摄，因而可能对于任意光源位置和任意拍摄角度的图片表现欠佳。使用不同照明位置或拍摄位置的数据进行训练可能会有所帮助。

**5.4. Extension and future work**

你会建议作者做哪些额外的实验来强化结果?

你能想到论文中提出的方法/想法(假设有效)的其他可能应用吗?

未来可能的作品是什么?

作者可以与其他方法在处理时间上做对比，从而验证方法是否time consuming.

他们还可以收集更多的从不同位置拍摄的以及不同光源位置的数据进行训练以适应更宽泛的拍摄条件，或者通过估计光源位置和反射率来辅助估计法线，从而优化表现。

未来该方法可能会用于游戏建模、影视等领域，并且入门门槛会更低。

**5.5. Is the paper stimulating or inspiring ?**

许多论文(甚至那些已经发表的论文)都是枯燥乏味的，而有些则是令人兴奋的。你对这篇论文有什么看法?为什么?

这篇论文是令人兴奋的。它解决了先前的方法需要密集输入或严格的外部条件的限制，使得通过图像进行人脸形状建模更容易实现，并且可以得到高质量的结果。

**5.6. Conclusion and personal reflection**

首先，对本文进行总结。

那么，如果让你来解决这个研究问题，你会有什么不同的做法?还有别的解决方案吗?

最后，用一句话总结你从阅读本文中学到的东西。

In conclusion，这篇论文提出了一个轻量级的两阶段卷积神经网络，把基于深度学习的光度立体技术与人脸先验信息结合，可以在近场点光源照明条件下，用稀疏的输入进行高精度的三维人脸建模，解决了其他方法需要密集输入或光照条件难以满足的缺陷。同时他们还扩充了相关的数据集。

如果是我，我可能会尝试在该方法的基础上根据【2】加入对光源位置和反射率的估计以辅助确定法线。这可能会减慢速度，但是应该可以提高准确度和对不同拍摄角度和光照角度的适应性。

在实验思想上，将神经网络与先验信息结合是这篇论文给我的最大的inspiration。