北京大學

本科生毕业论文

题目：三维模型最优二维视图生成方法研究

Synthesizing best 2D views of 3D models

姓　　名：黄道吉

学　　号：1600017857

院　　系：元培学院

本科专业：计算机科学与技术

指导老师：连宙辉

二〇二〇　年　五　月

摘要

生成三维模型的最优视图任务要求给定一个三维模型，我们能够选取出合适的视角，并且在这个视角下渲染出具有真实性的图片。随着三维建模方法的不断完善，三维模型的使用在近几年正变的越来越广泛。大规模三维模型数据库的出现更加便捷了有关三维模型的研究的进展。而三维模型不同于二维图片的特性使得它需要经过渲染才能显示在屏幕上。生成符合人类感知的三维模型的最优视图，将大大方便三维模型库的检索。近年来神经网络在二维视图生成工作的进展飞速，这也使得借助神经网络来生成三维模型的最优视图成为可能，尤其是生成对抗网络和变分自编码器在条件生成和非条件生成领域取得了很好的效果。本文回顾了以往在这个方向上研究者的工作，包括最优视角选择和新视角生成的算法，并提出了新的生成三维模型视图的方法。我们的方法首先能够根据一张导引图片提供的材质信息来渲染给定的三维模型在某一个视角下的视图，也能够通过在隐空间中采样来无条件的生成多样的三维模型的视图。定性和定量的实验结果表明，我们的新算法能够产生更加准确、更具真实性的三维模型图像，也在选择视角方面相对其他算法有自己的优势。最后，我们总结了本文的成果并展望了未来的工作。

关键词: 最优视角选择 新视角生成 材质迁移

Abstract

Synthesizing the best view of 3D models aims to choose to best viewpoint and rendering a realistic image given a 3D model. With the development of 3D modeling, the use of 3D models has been more and more prevalent. The emergent of 3D databases further advocates the development of 3D model related research. Unlike 2D images, 3D models have to be rendered to be displayed on a screen. Synthesizing the best view of 3D models in line with human perception will contribute to the retrieval of 3D databases. Thanks to the rapid development of neural networks' application on synthesizing 2D images, it has been possible in recent years to synthesize the best view of 3D models using neural networks, especially GAN and VAE, which yields great results in conditional and unconditional image synthesis. This paper reviews past research on this field, including best view synthesis and novel view synthesis, and propose a novel method to synthesize views of 3D models. First, our model can render an image of a given 3D model under a specific viewpoint while being consistent in style to a reference image. Second, our method can render various views of 3D models by sampling in latent space. Our new method is proved to be able to yield more accurate and realistic images through qualitative and quantitative experiments and outperforms previous best view selection methods. Finally, we summarize our contribution and proposed future work on this topic.

Key Words: best view selection novel view synthesis style transfer

全文目录

第一章 引言

1. 研究背景

三维模型是图形学和计算机视觉方向的研究重点。近年来，三维模型的应用变得越来越广泛，从游戏界和工业界的 CAD 模型，到前沿领域的自动驾驶，使用三维模型正大大便利着业界。RGB-D 传感器的应用也使得产生三维模型更加容易。在学术界，三维模型也有着广泛的应用：三维模型的分割 ([1, 2])、重建 ([3, 4])，以及利用三维模型强化对图片的理解 ([5])。这些因素都催生了大规模三维模型库的产生和广泛使用 (如 Shapenet [6], Pascal3D+ [7], ModelNet [8])。

在如此多的精力投入利用数据集解决问题的同时，相对少的精力投入到利用数据驱动的方法方便数据集的可视化和检索上。不同于二维图片便于观看、容易生成缩略图，三维模型在不同视角下会有不同的姿态，并且需要材质信息才能渲染出一张图片。这使得检索三维模型的数据库是一件费事的工作。ShapeNet 数据集 [6] 将每一个类别的模型对齐到同一个朝向，并在固定的方向渲染了 8 张缩略图，ModelNet 数据集 [8] 只提供了三维模型，这些方法并不能提供一个便捷的检索三维模型的方案。现有的处理三维模型的软件 (如 MeshLab [9])，提供用户一个拖拽视角的界面，让用户寻找最好的视角。如果能设计出生成最优视图的算法，将会便利检索三维模型数据库。

我们认为生成三维模型的最优视图至少包括两个部分，一个部分是选定最优的视角，另一部分是在这个选定的视角下渲染出带有材质的二位视图。第一个部分以往工作主要从图形学入手，通过在三维模型的顶点或是在二维视图上定义信息 (熵)，取熵最大的视角作为最优视角。渲染材质的工作则集中利用了基于神经网络的生成模型，将材质生成问题定义为有条件的图片到图片翻译的问题。我们借鉴了这两方面方法的核心思想，并提出了新的生成三维模型最优视图的算法。

2. 相关工作

2.1 最优视角选择

三维模型的最优视角选择任务旨在对给定的三维模型给出符合人类认知的最优的视角。这并不是一个良定义的问题，以往的研究方向往往采用在三维顶点或是二维像素上定义某种函数而将其转换为最优化问题。传统上认为最佳视角是包含最多信息的视角，不同的方法对信息的定义各不相同。在三维模型的二维视图上定义信息的文章主要包括：视角熵 [10]，曲率熵 [11]，轮廓熵 [11]，在不同的视角的投影中取信息最大的投影作为最佳视角。[12] 文中对比了几种基于几何学的方法的结果和人为标注的最优视角的差别，文章得出 MeshSaliency [13] 和视角熵 [10] 的方法是效果最好的传统方法。Mesh Saliency [13] 通过在每一个三维定点定义与曲率有关的显著性，并将可见的显著性加和最大的视角定义为最佳视角。文章更加提出了一种在视角空间中类似梯度下降的方法寻找最优视角的方法，而不需在视角空间中方格搜索 (grid search) 最优视角。视角熵 [10] 的方法关注二维投影中可见的每一个三维面片的投影面积，并将投影面积构成的分布的熵最大的视角作为最优的视角。我们认为这些方法有时并不会产生令人满意的效果。他们破坏了同一类三维模型共享同一个最佳视角的规则，并且对三维模型的建模方式很敏感。本研究首先复现了经典的传统方法，在以后的行文中，采用 Mesh Saliency 和视角熵作为传统方法的代表，和我们提出的方法作比较。

2.2 新视角生成

新视角生成 (novel view synthesis) 旨在给定一个三维模型在一个或多个视角下的视图来生成新视角下的视图。因为不同视角下可见的像素不同，这个任务本质上是一个非良定义的问题，而需要足够强的先验知识和正则化约束来得到可接受的结果。以往解决新视角生成任务的方法大致可以分为两大类：基于几何学的和基于学习的方法。几何学的方法能够从输入图片显式的估计三维模型的结构和材质信息。Multi-view stereo [14] 方法可以通过多个视角的输入图片直接重构出三维模型。Flynn et al. [15] 提出的深度神经网络能够在不同视角的图片中进行插值。几何学的方法主要缺点是作为训练数据的三维模型难以获得，并且缺失的像素会导致错误的破洞填补 (hole-filling)。基于学习的方法将新视角生成看作

图片生成任务，或采取预测从原图片到目标图片的流 [16–18] 的方式，或是采用某种正则化后直接生成每一个像素 [19–22]。不同的方法针对它非良定义的特性使用了不同的正则化方法，如感知损失函数 [18]，生成对抗网络的损失函数 [20] 和三维信息 [21] 的方式。本文提出的模型可以看作 [21] 的进一步拓展，将文中的三维信息进一步拆分为材质和内容信息，从而能够应用到材质迁移任务上。本文的方法在应用到新视角生成时，与流方法和像素生成的经典方法比较均有定性和定量结果的优势。

2.3 材质迁移

材质迁移旨在给定一张内容图片和一张材质图片 (或材质属性)，生成一张具有前者内容和后者的材质信息的图片。一种通用的思路是从输入图片中编码出表征内容和表征材质的向量，如 [23–26]。囿于成对训练数据缺失，材质迁移方法无法使用有监督的一范数或二范数损失函数，而需要采取独特的方法训练内容和材质的分离：如认为内容与材质信息分出 VGG 的不同层中 [27]，最小化循环损失 (cycle loss) [28] 或增加独特的判别器判定材质信息是否一致 [29, 30]。受 AdaIN [31] 启发，另一种融合材质和内容的方法是从内容或材质信息中回归参数来调整神经网络中间层激活量的大小和偏移，如 [32, 33]。在三维领域，将一个三维模型的材质迁移至另一三维模型或直接生成三维模型的材质，可以通过直接在三维空间中生成每一个点或面片的颜色 [34]，也可以将三维模型投影至二维空间进行，如 VON [35]。前者的方法生成结果较为模糊，并且因渲染过程不可导，需要采用近似渲染方法和可微分的渲染器 [36]。而后者能够借用图像领域材质迁移的成熟方法，取得更好的效果，并且作为这些方法中常用的深度图，也能够很好的在二维空间表征三维的信息。因此本文采用类似 VON 的模式，将三维模型渲染为深度图后，在深度图上渲染模型的材质信息。这个材质信息可以来自材质图片，也可以是真实图片，或者是随机采样的隐变量。定性和定量的结果表明我们的方法在生成图片质量上又一定的优势。

3. 研究内容

本文研究生成三维模型的最优二维视图的任务。它要求对给定的三维模型我们能够生成符合人类观感的图片，这包括选取出一个合适的视角和生成具有真实性的材质两部分，其中材质可以来自参考图片也可以非条件的生成。我们回

顾了在这个任务上前人的工作，综述并且实现了在视角选择和新视角生成领域经典的算法。在前人工作的基础上，我们提出了新的模型，将以往方法中的隐空间一分为二，分别编码三维模型的与视角无关的内容和材质信息，利用重构损失和对抗损失函数训练隐变量到图像空间的映射。我们认为最优视角是模型包含最多信息的视角，这种信息可以通过三维模型的某一视角下的图片重构不同视角下视图的精确度来衡量。在实验结果上，我们将我们的结果和复现的经典方法作比较，在定性结果和定量结果上，均体现出了优势。

4. 本文结构

第一章引言介绍了本文的研究背景，在回顾了相关工作的基础上阐述了本文的研究内容。第二章将会介绍本文的实验原理，包括生成对抗网络 (GAN) 和变分自编码器 (VAE) 的推导，以及应用在本文的情境下的公式。其后，第三章将介绍本文提出的模型结构。我们的方法和其他方法的结果比较将放在第四章中。最后，我们在第五章总结本文的工作，并且展望未来可能的进展方向。

参考文献

[1] Xiaobai Chen, Aleksey Golovinskiy, and Thomas A. Funkhouser. A benchmark for 3d mesh segmentation. In *SIGGRAPH 2009*, 2009.

[2] Abhijit Kundu, Yin Li, Frank Dellaert, Fuxin Li, and James M. Rehg. Joint semantic segmentation and 3d reconstruction from monocular video. In *ECCV*, 2014.

[3] Christopher Bongsoo Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. *ArXiv*, abs/1604.00449, 2016.

[4] Priyanka Mandikal, L. NavaneetK., Mayank Agarwal, and Venkatesh Babu Radhakrishnan. 3d-lmnet: Latent embedding matching for accurate and diverse 3d point cloud reconstruction from a single image. In *BMVC*, 2018.

[5] Christopher Bongsoo Choy, Michael Stark, Sam Corbett-Davies, and Silvio Savarese. Enriching object detection with 2d-3d registration and continuous viewpoint estimation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2512–2520, 2015.

[6] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *ArXiv*, abs/1512.03012, 2015.

[7] Yu Xiang, Roozbeh Mottaghi, and Silvio Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. *IEEE Winter Conference on Applications of Computer Vision*, pages 75–82, 2014.

[8] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2014.

[9] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, 2008.

[10] Pere-Pau Vázquez, Miquel Feixas, Mateu Sbert, and Wolfgang Heidrich. Automatic view selection using viewpoint entropy and its applications to image-based modelling. *Comput. Graph. Forum*, 22:689–700, 2003.

[11] David L. Page, Andreas F. Koschan, Sreenivas R. Sukumar, Besma Roui-Abidi, and Mongi A. Abidi. Shape analysis algorithm based on information theory. *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, 1:I–229, 2003.

[12] Helin Dutagaci, Chun Pan Cheung, and Afzal Godil. A benchmark for best view selection of 3d objects. In *3DOR@MM*, 2010.

[13] Chang Ha Lee, Amitabh Varshney, and David W. Jacobs. Mesh saliency. In *SIGGRAPH 2005*, 2005.

[14] Yasutaka Furukawa and Carlos Hernández. Multi-view stereo: A tutorial. *Found. Trends. Comput. Graph. Vis.*, 9(1-2):1–148, June 2015.

[15] J. Flynn, I. Neulander, J. Philbin, and N. Snavely. Deep stereo: Learning to predict new views from the world's imagery. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5515–5524, June 2016.

[16] Tinghui Zhou, Shubham Tulsiani, Weilun Sun, Jagannath Malik, and Alexei A. Efros. View synthesis by appearance flow. In *ECCV*, 2016.

[17] Shao-Hua Sun, Minyoung Huh, Yuan-Hong Liao, Ning Zhang, and Joseph J Lim. Multi-view to novel view: Synthesizing novel views with self-learned confidence. In *European Conference on Computer Vision*, 2018.

[18] Kyle Olszewski, Sergey Tulyakov, Oliver Woodford, Hao Li, and Linjie Luo. Transformable bottleneck networks. *The IEEE International Conference on Computer Vision (ICCV)*, Nov 2019.

[19] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Multi-view 3d models from single images with a convolutional network. In *European Conference on Computer Vision (ECCV)*, 2016.

[20] Rui Huang, Shu Zhang, Tianyu Li, and Ran He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[21] Xiaogang Xu, Ying-Cong Chen, and Jiaya Jia. View independent generative adversarial network for novel view synthesis. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[22] Eunbyung Park, Jimei Yang, Ersin Yumer, Duygu Ceylan, and Alexander C. Berg. Transformation-grounded image generation network for novel 3d view synthesis. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 702–711, 2017.

[23] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[24] Hung-Jen Chen, Ka-Ming Hui, Szu-Yu Wang, Li-Wu Tsao, Hong-Han Shuai, and Wen-Huang Cheng. Beautyglow: On-demand makeup transfer framework with reversible generative network. pages 10034–10042, 06 2019.

[25] TingTing Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *MM '18*, 2018.

[26] Wayne Wu, Kaidi Cao, Cheng Li, Chen Qian, and Chen Change Loy. Disentangling content and style via unsupervised geometry distillation. *ArXiv*, abs/1905.04538, 2019.

[27] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.

[28] Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. Conditional cyclegan for attribute guided face image generation. *ArXiv*, abs/1705.09966, 2017.

[29] Miao Wang, Guo-Ye Yang, Ruilong Li, Run-Ze Liang, Song-Hai Zhang, Peter. M Hall, and Shi-Min Hu. Example-guided style-consistent image synthesis from semantic labeling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[30] Liqian Ma, Xu Jia, Qianru Sun, Bernt Schiele, Tinne Tuytelaars, and Luc Van Gool. Pose guided person image generation. In *Advances in Neural Information Processing Systems*, pages 405–415, 2017.

[31] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017.

[32] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization, 2019.

[33] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[34] Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *ECCV*, 2018.

[35] Jun-Yan Zhu, Zhoutong Zhang, Chengkai Zhang, Jiajun Wu, Antonio Torralba, Joshua B. Tenenbaum, and William T. Freeman. Visual object networks: Image generation with disentangled 3D representations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[36] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

本科期间的主要工作和成果

本科期间参加的主要科研项目
本研基金

1. 国家创新训练计划. 基金类型. 连宙辉. 2018-2019

致谢

感谢