CHAPTER 8. CONCLUSIONS

Throughout this thesis, we have addressed important challenges in the fields of digital materials and radiance encoding, including *Visual Attribute Transfer* (Chapter 2), *SVBSDF Capture* (Chapter 2.C), *Tileable Texture Synthesis* (Chapter 3), *Neural BTF Propagation and Encoding* (Chapter 4), *SVBRDF Estimation* (Chapter 5), *Fabric Mechanical Parameter Capture* (Chapter 6) and *Neural Methods for Global Illumination* (Chapter 7). In this final section, our goal is to provide a comprehensive summary of the contributions made in this thesis and their significance in the fields of computer vision, graphics, and machine learning. By providing a broader view, we aim to highlight the importance and impact of the work presented here, both in terms of advancing the state-of-the-art and in its potential for real-world applications.

Digital Material Generation

The first projects developed in the context of this thesis are concerned with generating high quality digital materials, which can then be used for other tasks, like training neural material representations or creating large datasets for material estimation models. These methods take as input a single material, which helps increase the predictability of their outputs.

In Chapter 2, we present a neural visual attribute transfer framework capable of transferring, for a given material, many types of visual property maps to images of new patches of the same (or similar) material. These novel images may be taken under any illumination conditions, color variations, affine distortions, or camera setup. This model uses lightweight image-to-image convolutional neural networks to provide the first solution capable of leveraging the optical behavior of the material for the visual attribute transfer problem, by being trained on with a photometric dataset and a comprehensive data augmentation policy. This method can be trained in less than one minute, and is capable of transferring any type of visual attribute, from surface normals to stylizations or semantic segmentation maps. However, it is limited in accuracy, can only learn from one dataset at once, and can only learn to transfer one visual attribute.

To address these shortcomings, we extend this method in Chapter 2.C, where we propose a neural network capable of generating high-resolution large-scale SVBSDF maps of any material given one or more photometric datasets which represent small portions of a material. By extensive changes to the training procedure and improvements to the model architecture, not only do we enable this visual attribute transfer with the capability to learn from multiple datasets and transfer multiple images simultaneously, but we also increase the accuracy and sharpness of its estimations. However, these improvements increase the computational cost of training the neural network, and limit their generality. With these models, we can build large-scale materials, which can be used to generate BTF data to train other representations (Chapter 4) or build large datasets to train models that generalize to new materials (Chapter 5).

Figure 8.1: Renders generated with SEDDI's Textura.ai materials, estimated with methods derived from some of the contents in this thesis.

Textures used in rendering applications also need to be tileable. That is, when they are spatially concatenated, the seams between tiles are not visible. In Chapter 3, we introduce a single-image Generative Adversarial Network (GAN) which can generate multiple tileable textures of the same material. Our method allows for tileable outputs using a novel latent space manipulation algorithm, and handles multiple texture maps simultaneously, through neural architecture improvements. Further, we leverage the GAN's discriminator as a test-time quality estimation function, which helps distinguish between high quality and low quality outputs. This model achieves higher quality outputs than previous work, at lower computational times.

These models rely on a single material as input, which, while increasing predictability, limits their generalization capabilities. Nonetheless, our findings provide evidence that under certain assumptions and highly controlled training procedures, neural networks can be utilized for generating ground truth data. This data can be leveraged for other tasks, as demonstrated in other sections of this thesis.

Neural Radiance Encoding

One of the research problems that we explore in this thesis is the challenge of representing radiance using neural networks, which offers several potential benefits compared to traditional approaches. Neural networks can provide efficient, continuous, and fully differentiable functions that are useful for forward and inverse rendering. Additionally, they enable novel editing capabilities and efficient sampling for Monte Carlo rendering. In this work, we focus on two key components of virtual scenes: materials and illumination. Our goal is to develop neural network-based approaches for encoding these elements that are both accurate and computationally efficient, paving the way for new applications in computer graphics and vision.

For materials, in Chapter 4, we introduce a novel neural representation for encoding reflectance. Building upon previous work on the topic of neural materials [Kuz+21], our method has three components: A *neural texture*, which encodes the reflectance of the material in a 2D latent space; a fully-convolutional *neural renderer*, which takes as input this latent reflectance and light and view angles and outputs linear RGB values; and an *autoencoder*, which can estimate the neural texture of any input 2D image of the material. Similarly to previous work on neural fields [Mil+20], we train each model from scratch for every material, taking as input its *Bidirectional Texture Function* (BTF), which can be either captured or synthetically generated. Our approach introduces the first neural BTF representation with conditional input, which makes it possible to extrapolate BTF measurements, create tileable BTFs, and synthesize new materials through reflectance propagation. We address one of the main limitations of prior work on neural materials, which was their lack of editing capabilities, hindering their applicability to real-world rendering scenarios that require tileable and large-scale materials. Despite these improvements, our representation has limitations in that semantic properties such as specularity or albedo are encoded on the latent space and thus not easily editable. However, we believe that our method has significant potential for future research on neural material representations.

In Chapter 7, we addressed the problem of illumination encoding, introducing a novel neural method for joint sampling, PDF encoding, and RGB compression of environment maps used for global illumination in rendering. Our method builds on previous work on *implicit neural representations* for compressing the HDRi environment map into a continuous, differentiable function. We designed a *Normalizing Flow* that can sample directions from environment maps and measure their probability density, both of which are essential for Multiple Importance Sampling in Monte Carlo rendering. Our carefully designed neural architecture and training procedures allowed us to obtain models that are up to two orders of magnitude faster than analytical methods while still being fully differentiable. Using a dataset of environment maps with diverse properties, we demonstrated high generality and accuracy, outperforming previous work on the topic. To encourage future research, we will provide an open-source implementation and a dataset of trained models. While this is the first method that learns to sample from environment maps, we believe there is potential to further improve our approach by introducing more sophisticated neural network designs or learning priors over global illumination.

Despite their shortcomings, these two neural methods introduce contributions to the field of neural scene representations, and they could both be incorporated into neural rendering pipelines and more traditional path-tracing engines. We hope these ideas inspire further research interest in this field.

Scalable Material Digitization

One of the key objectives of this thesis is to develop digitization solutions that are both scalable and affordable, without relying on expensive or inaccessible hardware. To accomplish this, we have made two different contributions. First, in Chapter 5, we introduce a novel method for estimating *spatially varying bidirectional reflectance distribution functions* (SVBRDFs) using flatbed scanners as the capture device. This method strongly relies on the data generation algorithms presented in previous sections of this thesis. Our approach employs a custom-built, lightweight neural network with attention mechanisms, resulting in estimations with considerably higher resolution than previous methods that relied on smartphone flash-lit images as input. We empirically observe that material reflectance is strongly dependent on its microgeometry, which flatbed scanners can capture appropriately, all the while providing images that can be directly used as the albedo of the SVBRDF. One of the key contributions of this work is the introduction of the first uncertainty quantification algorithm for material estimation, for which we demonstrate applications on dataset creation through active learning. Further, an extension of this work is part of Textura.ai, which is utilized by real-world users, providing evidence of its robustness and reliability. In Figure 8.1, we show some digital garments which use materials digitized with this technology. We are confident that the novelties introduced by this method will shape future research on the long-standing problem of single image material digitization. Although we are confident in the method's capabilities, it can be extended to estimate additional reflectance properties like anisotropy or transmittance, to generalize to other materials or devices, and to improve its accuracy even further.

Besides, in Chapter 6, we proposed a method for mechanical parameter estimation of textile materials using depth images and the material density as inputs. Our method achieves greater scalability than previous work, which required video sequences, manual intervention, or purposefully built and expensive devices. To do so, we propose a solution that is agnostic to the optical appearance of the fabric, and can be utilized by non-expert operators. Our custom-built multi-input neural network predicts the full set of mechanical parameters of any textile using synthetic data, leveraging a strong data augmentation policy, pre-training, attention mechanisms, and multiple forms of pooling of internal activations. A significant contribution of this work is our proposed *drape similarity metric*, which measures distances between the mechanical behavior of fabrics while using only images as input. We validate this metric with a user study conducted on-site, showing that it strongly correlates with human perception. Although our method has significant contributions to estimating the mechanical behavior of textiles, it is still limited in at least two ways: it takes depth images as input, which are relatively easy to obtain but are less accessible than RGB pictures, and the proposed drape similarity metric is not easily differentiable, hindering its potential as a loss function in optimization problems.

These two estimation algorithms, along with the uncertainty quantification algorithm and the proposed drape similarity metric, have the potential to transform the low cost material digitization field, introducing more scalable, robust, and perceptually-validated solutions.

New Frontiers

Although this thesis has made significant contributions to the fields of radiance encoding, and digital material capture and edition, we acknowledge that our proposed methods have certain limitations in terms of scope and generality. Additionally, concurrent developments in deep learning, neural rendering, and computer vision have opened up new and exciting research avenues that will shape the future of the field. With this in mind, we would like to highlight several research directions that we believe are particularly intriguing for extending the work presented in this thesis.

Generative Models for Materials

In this thesis, we have presented models that are limited to solving individual tasks such as material digitization, encoding, or tileable texture synthesis. However, recent research [Rom+22; HJA20; Sah+22] has demonstrated that a single generative model can be used to solve multiple tasks simultaneously. Developing a single generative model for materials that can be used for capture, super-resolution, edition, real-time tileable synthesis, or material interpolation would be an exciting research direction for the future of digital materials. One of the main challenges in this area is the lack of publicly available datasets of real materials, which are essential for reproducible research on deep learning algorithms. Nevertheless, techniques like pre-training, fine-tuning, and data augmentation policies may be employed to construct these models without requiring massive datasets of materials.

Multimodal Learning

The findings in this thesis suggest an interesting research avenue: multimodal learning for materials. In addition to capturing a material's reflectance and mechanical properties, it is possible to digitize other attributes such as its fabrication process, semantic attributes, fine-grained composition, origin, cost, or even its carbon footprint. Recent work on multimodal and contrastive learning [Par+20] suggests that a shared embedding for visual and semantic attributes can be found, which can then be utilized for various tasks such as zero-shot learning or conditioning generative models [Rom+22]. Developing such an embedding for materials could be extremely valuable for a wide range of problems, including material edition, interpolation, classification, generation, similarity measurement, or anomaly detection. This research direction is promising, but it also poses significant challenges, such as the need for large and diverse datasets that include multiple modalities of material information, and the development of effective multimodal learning architectures that can handle such information. Nonetheless, we believe that the benefits of multimodal learning for materials make it a hugely compelling direction for future research.

Neural Spatially-Varying BSDFs

In Chapter 7, we demonstrated the effectiveness of invertible generative neural networks for efficient Multiple Importance Sampling in global illumination. Similar networks have also been utilized for BRDFs [Szt+21], suggesting that learned sampling could be applied to spatially-varying materials. Further, in Chapter 4, we propose a neural representation of spatially-varying materials, which showed promising compression results. Combining the ideas of these two projects, it should be possible to train invertible neural networks to learn to sample from real-world materials, which often exhibit complex optical behavior such as spatially-varying reflectance or transmittance, we can potentially improve the efficiency and quality of path-traced renders while enabling new inverse rendering applications. This presents a promising direction for future research in material capture and rendering.

Uncertainty Quantification and Model Understanding

We have presented the first method for uncertainty quantification in material capture applications, demonstrating its utility for active learning and dataset creation. Additionally, we have shown that visualization of material capture models can aid in understanding their predictions. However, we believe that these areas are still relatively unexplored, and further analysis could provide valuable insights for designing the next generation of material digitization pipelines. This includes developing more suitable capture devices, identifying valuable data points for labeling, and building more accurate models. Furthermore, we believe that investigating methods for communicating to end users when a capture may be inaccurate and suggesting alternative digitization pathways is an important research topic that could enhance the impact, robustness, and trustworthiness of these models in the industry.

Enhancing Digitization Quality and Realism

In this thesis, we have presented methods for highly realistic material encodings (Chapter 4) and digitization pipelines (Chapter 2.C). They can be used to represent a wide variety of materials at very high qualities, however, they require expensive capture systems. On the other hand, in Chapter 5, we introduce a single-image digitization system capable of capturing a small set of SVBRDF parameters. While the model is adequately accurate, this limited material model cannot represent complex optical phenomena, like anisotropy or transmittance. Extending single-image material digitization systems to be able to accurately estimate more reflectance properties is a very challenging research problem, because this estimation is inherently ill-posed, and there is very limited available training data. These low cost capture systems will be able to increase the realism of their digitizations in the future, as more and better data becomes available, and better models can be built as the machine learning field advances.

Enhancing Generality on Devices and Materials

A limitation of our casual capture algorithms is that they require the use of specific, albeit low-cost, devices such as flatbed scanners or depth cameras. An immediate extension of our methods is to make them compatible with other types of devices, particularly smartphone cameras, which would increase their applicability. However, less controlled capture settings typically hinder the accuracy of digitizations, as additional assumptions must be made to obtain a plausible result. With the ongoing improvement in the quality of smartphone cameras, along with larger datasets and better deep learning models, high-quality digitization of material reflectance or fabric mechanics using smartphones could become a reality in the near future.

In addition, like any learning-based system, our models are constrained by the materials present in their training set, limiting their ability to generalize to other materials. Although we have achieved accurate digitization of fabrics and leathers, there is no guarantee that our models will perform well on out-of-distribution materials such as stone, plastics, metals, or glass. While some of these materials may be successfully digitized using single-image models with sufficient data, others may present more significant challenges. It is our hope that future research in this area will help improve the generality and accuracy of these models.

Digitization on Edge Devices

One limitation of material capture algorithms based on deep learning is that they require expensive hardware for inference. This means that in commercial products, users can capture images on a commodity device, but the digitization process is typically carried out on a cloud-based application. As a result, bottlenecks may arise that hinder the efficiency of the digitization process, with users having to wait several minutes before obtaining their digital material. To achieve real-time learning-based digitization, it may be necessary to perform model inference on edge devices such as smartphones, tablets, or laptops. However, this approach may place limitations on model sizes, potentially leading to a reduction in digitization accuracy. Recent work on edge deep learning [Wan+20c] may provide interesting cues for future research on this topic.

Neural Scene Representations

Concurrently to the work in this thesis, *Neural Radiance Fields* [Mil+20] emerged as an alternative powerful representation for virtual scenes. A Cambrian explosion of research in radiance fields for scene representations has emerged to increase their capabilities, in terms of data efficiency, speed, realism, or generality [Wan+22d; Mül+22; Wan+23; Che+22; Tew+22; Ver+22; Fri+22; Att+22; Xie+22]. However, these remain somewhat limited in controllable and capabilities for edition, for which more traditional rendering and scene creation pipelines may be more adequate than learning-based alternatives. Our neural encodings for illumination or materials, which may be integrated into either neural representations or path-traced rendering engines, may help bridge these two pipelines, which should benefit both.

Dataset Generation for Vision Downstream Tasks

Synthetic data has been extensively shown to help deep learning models generalize to real images [Azi+23; Zhu+21; Ric+16]. Our proposed method for casual capture of mechanical parameters in fabrics (Chapter 6, [Rod+23b]) provides additional evidence in this direction. The procedural creation of virtual environments for pre-training of computer vision models will remain a very relevant topic in future years in the field, as it will reduce the need for expensive labeling of real-world data. These virtual environments strongly benefit from highly accurate assets for enhanced realism, as well as efficient yet expressive image formation algorithms. We hope that the contributions in this thesis will help in generating synthetic data for training computer vision models which solve general tasks, including classification or segmentation [Kir+23].

Applications to Other Fields

Finally, a fascinating use of computer vision and material digitization is *digital cultural heritage* [HMV09]. Digital capture of artifacts from ancient civilizations could help important research fields, like anthropology or archaeology. However, past solutions required inaccessible devices [Ham+21; Dye+18], which limited their impact. Inexpensive gathering of photorealistic copies of ancient artifacts, which can be shared digitally, can provide valuable solutions for these research fields. We hope that low cost and accurate digitization systems may not only help the visual media and design industries, but also other fields for which digital asset inventories are helpful.

Closing Remarks

The collaborative efforts between disciplines and research institutions, both public and industrial, have been instrumental in enabling the contributions of this thesis, and the advancements made in the computer vision, graphics, and machine learning fields as a whole in the past few years. The interdisciplinary environment in which I worked during my PhD greatly enhanced the quality of my research and communication skills, while also providing me with the opportunity to learn fascinating new topics and collaborate with exceptional people. Further, while the massive amounts of publications in these fields put a significant amount of pressure on the researchers, it makes the field move incredibly quickly and creates a very exciting environment to work in. I hope that the focus on open science and the integration of ideas and feedback from other research areas continues, along with a greater awareness of the ethical and ecological impact of the models we create and deploy. Working as a researcher is in many ways a privilege, and it is our responsibility to advance science in a way that positively impacts society and the environment, and creates a better, more prosperous, sustainable, and equitable future for everyone. Looking ahead to the future, my hope is that the contributions made in this thesis will have a tangible impact on the real world and serve to further advance research in positive ways. On a personal level, my goal is to continue learning and challenging myself, all while striving to make a positive difference in the world.