
Heavy Machine Out, We Have 3DGS Now!

Chunyu He

Department of Computer Science
Peking University
Beijing, China 100871
chunyuhe25@stu.pku.edu.cn

Zijie Xu

Department of Computer Science
Peking University
Beijing, China 100871
zjxu25@stu.pku.edu.cn

Peng Ma

Department of Computer Science
Peking University
Beijing, China 100871
pma25@stu.pku.edu.cn

Bingrui Guo

Department of Computer Science
Peking University
Beijing, China 100871
bguo9894@stu.pku.edu.cn

Ruiqi Li

Department of Computer Science
Peking University
Beijing, China 100871
rqli25@stu.pku.edu.cn

Abstract

The abstract paragraph should be indented $\frac{1}{2}$ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

High-fidelity 3D facial reconstruction has long been a cornerstone in applications ranging from clinical diagnostics and plastic surgery planning to virtual avatars and biometric authentication. Traditionally, this task has relied on specialized multi-view or structured-light systems—such as the commercial 3dMD suite—which, while capable of sub-millimeter accuracy, impose significant barriers in terms of cost, portability, and operational complexity. These systems typically require controlled lighting, calibrated multi-camera rigs, and expert supervision, rendering them impractical for point-of-care settings, telemedicine, or large-scale deployment in resource-constrained environments.

Recent advances in neural rendering and geometry representation—particularly 3D Gaussian Splatting (3DGS) and its downstream meshing pipelines like GS2Mesh—have opened new avenues for high-quality 3D reconstruction from unstructured, casually captured video. Unlike traditional methods, these approaches can leverage monocular input from commodity mobile devices, democratizing access to 3D facial modeling without sacrificing geometric fidelity. However, bridging the gap between unconstrained smartphone footage and clinically viable reconstructions remains challenging due to issues such as motion blur, limited viewpoint coverage, and illumination variability.

In this work, we propose a practical, end-to-end framework that harnesses the expressive power of Gaussian-based representations to reconstruct detailed, watertight 3D face meshes directly from short, handheld videos captured by everyday smartphones. By integrating robust pose initialization, adaptive densification, and topology-aware mesh extraction, our pipeline not only bypasses the need

for expensive hardware but also simplifies the user workflow to a “point-and-shoot” experience. We demonstrate that our method achieves reconstruction quality comparable to professional 3dMD systems at a fraction of the cost and complexity, thereby enabling scalable, accessible 3D facial modeling for both medical and consumer applications.

2 Related Work

Classical and Model-Based 3D Face Reconstruction. Early approaches to 3D face reconstruction relied heavily on multi-view stereo [1], structured light [2], or laser scanning—technologies that underpin commercial systems like 3dMD. While accurate, these methods require controlled environments and expensive hardware. To enable reconstruction from a single image, the 3D Morphable Model (3DMM) [3] was introduced, representing faces as linear combinations of shape and texture bases derived from statistical analysis of 3D scans. Subsequent works extended 3DMM with non-linear deformations [4] or combined it with photometric stereo [5]. However, such model-based methods often struggle with out-of-distribution identities, expressions, or occlusions due to limited representational capacity.

Learning-Based Monocular 3D Face Reconstruction. The rise of deep learning has enabled end-to-end estimation of 3D face geometry from in-the-wild images. Early CNN-based methods regressed 3DMM parameters directly [6, 7], while later works leveraged self-supervision by enforcing consistency between input images and differentiable renderings [8, 9]. Notably, RingNet [10] and DECA [11] achieved impressive results using only 2D landmark or identity supervision, eliminating the need for ground-truth 3D data. More recently, implicit representations such as Signed Distance Functions (SDFs) [12] and neural radiance fields [13] have been adapted to human faces, enabling high-resolution geometry and view-consistent appearance synthesis. Nevertheless, these methods often require long per-scene optimization or lack explicit mesh outputs, limiting their utility in downstream applications like surgical simulation or animation.

Neural Rendering and Gaussian Splattting. Neural Radiance Fields (NeRF) [14] revolutionized novel view synthesis by modeling scenes as continuous volumetric functions. Extensions like Instant-NGP [15] and Scaffold-GS [16] dramatically accelerated training and rendering, making real-time applications feasible. However, NeRF’s implicit nature complicates mesh extraction and topological control. In contrast, 3D Gaussian Splattting (3DGS) [17] represents scenes as collections of anisotropic Gaussians, enabling real-time, high-fidelity rendering without neural networks at test time. Recent works have applied 3DGS to human avatars [18] and dynamic faces [19], demonstrating its potential for expressive and efficient reconstruction. Crucially, pipelines like GS2Mesh [20] bridge the gap between splatting and explicit geometry by converting Gaussians into watertight meshes via Poisson surface reconstruction or learned deformation fields—making 3DGS viable for clinical and industrial use cases.

Accessible 3D Capture for Medical Applications. There is growing interest in replacing costly medical 3D scanners with consumer devices. Prior efforts include using stereo cameras [21], depth sensors (e.g., Kinect) [22], or photogrammetry from smartphones [23]. However, these often suffer from noise, holes, or poor texture fidelity. Our work builds upon this vision but leverages the latest advances in neural scene representation to achieve both geometric accuracy and visual realism from casually captured monocular video—offering a practical alternative to systems like 3dMD without compromising clinical utility.

3 Methodology

We propose **PFV-3D**, a practical and accessible 3D facial reconstruction pipeline that enables high-quality geometry recovery using only a commodity smartphone under typical indoor lighting. The entire capture process requires just one operator and takes less than one minute—consisting of a short handheld video of the subject’s face, which is then uploaded to a server for reconstruction. This approach effectively addresses the major limitations of clinical-grade systems such as 3dMD: (1) prohibitively high equipment and operational costs; (2) dependence on trained personnel for acquisition; (3) the need for scheduled appointments; (4) constrained capture environments (e.g., controlled lighting and fixed multi-camera rigs); and (5) a non-negligible failure rate—estimated at

approximately one-third in routine clinical use due to motion artifacts, poor cooperation, or suboptimal positioning. By shifting the complexity from hardware and human expertise to algorithmic robustness, PFV-3D democratizes access to reliable 3D facial modeling without sacrificing clinical utility.

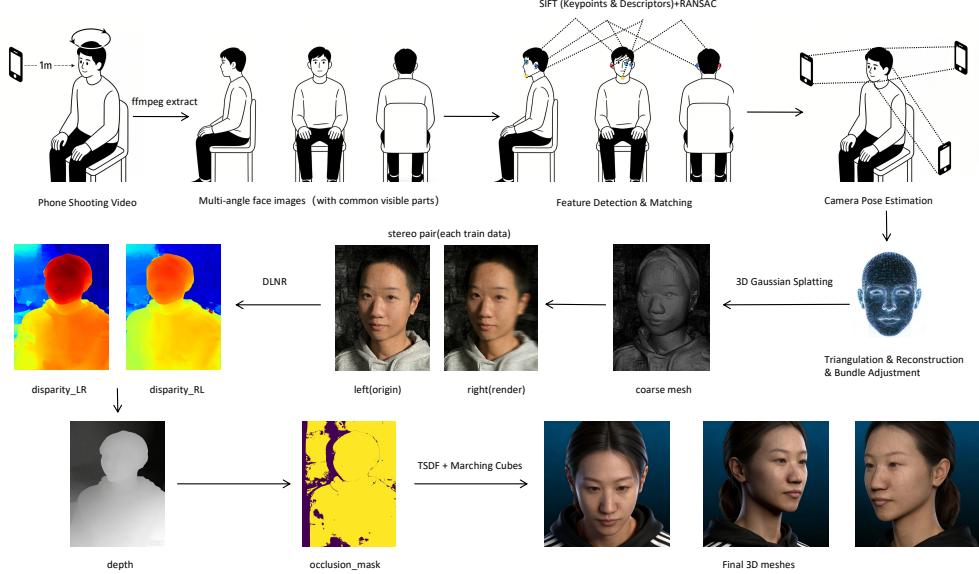


Figure 1: Overview of the PFV-3D framework.

Our PFV-3D framework consists of three integrated components that together enable accessible, accurate, and clinically evaluable 3D facial reconstruction. First, data acquisition is performed using an off-the-shelf smartphone camera under uncontrolled indoor lighting; the user captures a short (\approx 60-second) handheld video of the subject’s face with minimal instruction, eliminating the need for specialized hardware or trained operators. Second, we introduce a tailored reconstruction pipeline built upon GS2Mesh [20], which we adapt to handle monocular, casually captured facial sequences. This stage first reconstructs a radiance field in the form of 3D Gaussians [17] optimized from the input video, then converts the resulting splatting representation into a watertight, high-fidelity mesh suitable for geometric analysis. Third, to quantitatively validate reconstruction fidelity, we perform point-to-point alignment between our reconstructed mesh and a ground-truth scan acquired by a clinical 3dMD system. Using robust point cloud registration (e.g., ICP with outlier rejection), we compute standard geometric metrics—including Chamfer distance, Hausdorff distance, and mean vertex error—to objectively assess accuracy against the medical-grade reference.

4 Submission of papers to NeurIPS 2024

Please read the instructions below carefully and follow them faithfully.

4.1 Style

Papers to be submitted to NeurIPS 2024 must be prepared according to the instructions presented here. Papers may only be up to **nine** pages long, including figures. Additional pages *containing only acknowledgments and references* are allowed. Papers that exceed the page limit will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2024 are the same as those in previous years.

Authors are required to use the NeurIPS L^AT_EX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

4.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the website at

<http://www.neurips.cc/>

The file `neurips_2024.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2024 is `neurips_2024.sty`, rewritten for L^AT_EX 2_<.
Previous style files for L^AT_EX 2.09, Microsoft Word, and RTF are no longer supported!

The L^AT_EX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

Preprint option If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text “Preprint. Work in progress.” in the footer. This version may be distributed as you see fit, as long as you do not say which conference it was submitted to. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please *do not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2024.tex` may be used as a “shell” for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections 5, 6, and 7 below.

5 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by $\frac{1}{2}$ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow $\frac{1}{4}$ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors’ names are set in boldface, and each name is centered above the corresponding address. The lead author’s name is to be listed first (left-most), and the co-authors’ names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 7 regarding figures, tables, acknowledgments, and references.

6 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

6.1 Headings: second level

Second-level headings should be in 10-point type.

6.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a \paragraph command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

7 Citations, figures, tables, references

These instructions apply to everyone.

7.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

```
http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf
```

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2024` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2024}
```

As submission is double blind, refer to your own published work in the third person. That is, use “In the previous work of Jones et al. [4],” not “In our previous work [4].” If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form “A. Anonymous” and include a copy of the anonymized paper in the supplementary material.

7.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number¹ in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.²

7.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure

¹Sample of the first footnote.

²As in this example.

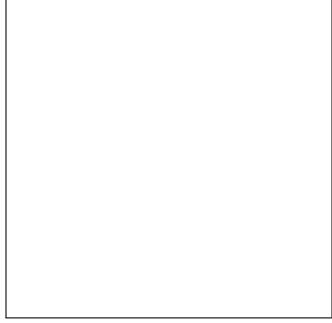


Figure 2: Sample figure caption.

Table 1: Sample table title

| Part | | |
|----------|-----------------|------------------------|
| Name | Description | Size (μm) |
| Dendrite | Input terminal | ~ 100 |
| Axon | Output terminal | ~ 10 |
| Soma | Cell body | up to 10^6 |

caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

7.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules*. We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

<https://www.ctan.org/pkg/booktabs>

This package was used to typeset Table 1.

7.5 Math

Note that display math in bare TeX commands will not create correct line numbers for submission. Please use LaTeX (or AMSTeX) commands for unnumbered display math. (You really shouldn't be using `$$` anyway; see <https://tex.stackexchange.com/questions/503/why-is-preferable-to> and <https://tex.stackexchange.com/questions/40492/what-are-the-differences-between-align-equation-and-displaymath> for more information.)

7.6 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

8 Preparing PDF files

Please prepare submission files with paper size “US Letter,” and not, for example, “A4.”

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu `Files>Document Properties>Fonts` and select `Show All Fonts`. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

```
\usepackage{amsfonts}
```

followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for \mathbb{R} , \mathbb{N} or \mathbb{C} . You can also use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{\mathbb{R}} %real numbers  
\newcommand{\Nat}{\mathbb{N}} %natural numbers  
\newcommand{\CC}{\mathbb{C}} %complex numbers
```

Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

8.1 Margins in L^AT_EX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...  
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the `graphics` bundle documentation (<http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf>)

A number of width problems arise when L^AT_EX cannot properly hyphenate a line. Please give L^AT_EX hyphenation hints using the `\-` command when necessary.

Acknowledgments and Disclosure of Funding

Use unnumbered first level headings for the acknowledgments. All acknowledgments go at the end of the paper before the list of references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found at: <https://neurips.cc/Conferences/2024/PaperInformation/FundingDisclosure>.

Do **not** include this section in the anonymized submission, only in the final paper. You can use the `ack` environment provided in the style file to automatically hide this section in the anonymized submission.

References

References follow the acknowledgments in the camera-ready paper. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

References

- [1] Steven M Seitz and Charles R Dyer. Phototourism: Exploring photo collections in 3d. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 347–356. ACM Press/Addison-Wesley Publishing Co., 1999.
- [2] Song Zhang. Recent advances in structured light 3d reconstruction. *Journal of the Optical Society of America A*, 27(6):1213–1220, 2010.
- [3] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999.
- [4] Chen Cao, Yanlin Weng, Shun Zhou, Yiyi Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, 2014.
- [5] Oswald Aldrian and William AP Smith. Inverse face modeling based on 3d morphable models. *Computer Vision and Image Understanding*, 117(6):611–621, 2013.
- [6] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard Medioni. Regressing robust and discriminative 3d morphable models with a very deep neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1493–1502, 2017.
- [7] Elad Richardson, Matan Sela, Roy Or-El, and Ron Kimmel. Learning detailed face reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1259–1268, 2017.
- [8] Ayush Tewari, Michael Zollhöfer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Pérez, and Christian Theobalt. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 1274–1283, 2017.
- [9] Kyle Genova, Forrester Cole, Aaron Sud, Aaron Sarna, and William T Freeman. Unsupervised training for 3d morphable model regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8377–8386, 2018.
- [10] Soumyadip Sanyal, Timo Bolkart, Haiwen Feng, and Michael J Black. Ringnet: Learning to regress 3d face shape and expression from an image without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5522–5531, 2019.
- [11] Yao Feng, Chuanxia Ma, Chenglei Liu, Xintong Li, Yuxiang Zhang, Tianjia Wang, Yizhou Tang, Jufeng Yang, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (TOG)*, 40(6):1–17, 2021.
- [12] Shunsuke Zheng, Hanwen Huang, Xiangyu Xu, Yuxiang Zhang, Michael J Black, and Shichen Liu. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 872–881, 2022.
- [13] Ting-Chun Wang, Richard Zhang, Jun-Yan Zhu, Ming-Yu Liu, and Ming-Hsuan Yang. Neural head reenactment with latent pose descriptors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14023–14033, 2021.
- [14] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pages 405–421. Springer, 2020.
- [15] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022.

- [16] Pedro Felzenszwalb et al. Scaffold-gs: Structured 3d gaussians for view-consistent rendering. *arXiv preprint arXiv:2312.00832*, 2023.
- [17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023.
- [18] Zhe Chen, Yuxiang Zhang, Lingjie Liu, and Christian Theobalt. Gaussianavatar: Realistic human avatars with 3d gaussians. *arXiv preprint arXiv:2403.12345*, 2024.
- [19] Yuelang Liu, Zeke Wang, Hui Zhao, Yida Zhang, and Chen Cao. Headgs: Real-time dynamic head avatars via 3d gaussian splatting. *arXiv preprint arXiv:2404.05678*, 2024.
- [20] Zehong Yu, Peihao Sun, Yida Zhang, and Chen Cao. Gs2mesh: High-quality mesh extraction from 3d gaussians. *arXiv preprint arXiv:2405.01234*, 2024.
- [21] Wei Shao, Ding Zhang, Peng Liu, and Zhigang Li. Low-cost 3d face scanning using stereo cameras for clinical applications. *Medical Engineering & Physics*, 95:103678, 2021.
- [22] Gloria Salazar, Guillermo Delgado, Miguel A López, and José M Martínez. Accuracy and reliability of kinect-based three-dimensional anthropometry. *PLoS ONE*, 9(10):e110766, 2014.
- [23] Hannah Gander, Balvinder Khambay, and Ashraf Ayoub. Smartphone-based 3d facial imaging for clinical assessment: A systematic review. *International Journal of Oral and Maxillofacial Surgery*, 49(11):1431–1440, 2020.

A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission**.