

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ALIGN TECHNOLOGY, INC.,

Petitioner,

v.

DENTAL MONITORING SAS,

Patent Owner.

Case No. IPR2024-00052

U.S. Patent No. 11,049,248

**DECLARATION OF DR. JOHN MONGAN IN SUPPORT OF
PATENT OWNER'S RESPONSE**

TABLE OF CONTENTS

I.	SCOPE OF ENGAGEMENT	1
II.	SUMMARY OF OPINIONS.....	1
III.	BACKGROUND	3
A.	EDUCATION AND INDUSTRY EXPERIENCE	3
B.	COMPENSATION	8
C.	MATERIALS AND OTHER INFORMATION CONSIDERED	8
IV.	UNDERSTANDING OF THE LAW	9
A.	PERSON OF ORDINARY SKILL IN THE ART	9
B.	DATE OF INVENTION	10
C.	PRIOR ART.....	11
D.	CLAIM CONSTRUCTION.....	11
E.	PATENTABILITY	13
F.	OBVIOUSNESS.....	13
V.	THE '248 PATENT	18
A.	THE EFFECTIVE FILING DATE OF THE '248 PATENT.....	18
B.	PERSON OF ORDINARY SKILL IN THE ART.....	18
C.	BACKGROUND ON ALIGNER TREATMENT	20
D.	INTRODUCTION TO THE '248 PATENT	21
E.	THE STATE OF THE ART RELATING TO NEURAL NETWORKS	26
VI.	OVERVIEW OF THE ASSERTED ART	32
A.	KIM	32
B.	INVISALIGN.....	46
C.	SALAH	51
D.	MANINIS	67
VII.	CLAIM CONSTRUCTION.....	81

A.	CONSTRUCTIONS PROPOSED BY DR. FOROOSH	81
B.	ADDITIONAL CONSTRUCTIONS	87
VIII.	THE '248 PATENT WOULD NOT HAVE BEEN OBVIOUS IN VIEW OF KIM, INVISALIGN, SALAH, AND MANINIS	92
A.	DR. FOROOSH'S PROPOSED COMBINATION	92
B.	DR. FOROOSH'S PROPOSED COMBINATION IS INOPERABLE	104
C.	KIM, INVISALIGN, AND SALAH FAIL TO DISCLOSE OR RENDER OBVIOUS A METHOD OF ASSESSING ALIGNER FIT USING AN ALIGNER IMAGE, AND THERE IS NO MOTIVATION TO COMBINE THE REFERENCES AS PROPOSED.....	108
D.	KIM, INVISALIGN, SALAH, AND MANINIS FAIL TO DISCLOSE OR RENDER OBVIOUS THE CLAIMED ANALYSIS OF AN ANALYSIS IMAGE BY MEANS OF A DEEP LEARNING DEVICE TRAINED BY A LEARNING BASE, AND THERE IS NO MOTIVATION TO COMBINE THE REFERENCES AS PROPOSED.....	115
IX.	CONCLUSION	137

EXHIBIT LIST

Exhibit Number	Description
2002	Declaration of Dr. John Mongan, dated January 30, 2024
2008	May 7, 2020 Press Release of Petitioner, Align Technology, Inc.
2009	U.S. Pat. Pub. 2017/0325689 (Dental Monitoring's English Language Translation of the International Patent Application published as Salah (PCT Publication No. WO2016066651))
2010	J. Liu et al., <i>Machine learning in orthodontics: Challenges and perspectives</i> , 30(10) <i>Advances in Clinical and Experimental Medicine</i> 1065
2020	The PASCAL VOC 2012 Segmentation dataset (July 13, 2017 version available at https://web.archive.org/web/20170713000522/http://host.robots.ox.ac.uk/pascal/VOC/voc2012/#data)
2029	Updated Curriculum Vitae of John T. Mongan, M.D., Ph.D.
2041	O. Russakovsky et al., <i>ImageNet Large Scale Visual Recognition Challenge</i> , https://arxiv.org/pdf/1409.0575
2042	S. Xie & Z. Tu, <i>Holistically-Nested Edge Detection</i> , https://arxiv.org/pdf/1504.06375
2045	B. Lang, <i>NVIDIA's Next-Gen Titan X Has 11 TFLOPS Performance, 12GB of Memory, and Costs 1 Arm and 1 Leg</i> , Road TOVR (July 21, 2016)
2048 ¹	Deposition Transcript of Dr. Hassan Foroosh, Ph.D., in IPR2023-01369 (May 17, 2024)
2053	S. Frey, <i>A Quick Look at Dental Monitoring, The Ortho Cosmos</i> , http://theorthocosmos.com/quick-look-dental-monitoring/ (Apr. 17, 2017)

¹ I understand that Patent Owner is not yet in receipt of Dr. Foroosh's signature or the officer's certification and will coordinate with Petitioner to file an updated exhibit once complete. 42.53(f)(5)-(6).

2054	Excerpts from Transcript of Deposition of Hassan Foroosh, Ph. D, in IPR2024-00110 and IPR2024-00052 (June 28, 2024)
2064	Provisional Patent Application No. 62/417,985 to Carrier et al.

I. SCOPE OF ENGAGEMENT

1. I have been retained by counsel for Dental Monitoring SAS as an expert witness in the above-captioned proceeding. I have been asked to provide my opinions relating to the state of the art for the technology described in U.S. Patent No. 11,049,248 (“the ’248 patent”) (EX1001) and on matters relating to the patentability of the challenged claims of that patent, as well as to respond to the opinions of Dr. Hassan Foroosh (EX1002). The following is my written declaration on these issues.

II. SUMMARY OF OPINIONS

2. In summary, my opinions are as follows:

- Dr. Foroosh proposes that a POSA would have been motivated to train and use Maninis’s COB neural network to detect semantic contours (*i.e.*, contours labeled as associated with a particular object), but Maninis’s COB neural network could not have been trained and would not have performed in the proposed manner. Dr. Foroosh’s proposed combination is thus inoperable and does not disclose the limitations of the ’248 patent claims nor would a POSA have been motivated to make such a combination with a reasonable expectation of success.
- The ’248 patent discloses and claims a brand-new method of assessing aligner fit from simple cellphone images. Kim, Invisalign, and Salah fail to disclose

or render obvious a method of assessing aligner fit using an aligner image, and there would have been no motivation to combine the references as proposed to arrive at the claimed methods. Dr. Foroosh misreads each of Kim, Invisalign, and Salah, none of which disclose or provide a reason to capture images of a patient's teeth with aligners in a service position, let alone for the purpose of analyzing such an image to determine aligner fit.

- None of the references disclose analysis of an aligner image by means of a deep learning device trained by a learning base so as to determine a tooth or image attribute relating to the separation between an aligner and a tooth, nor would a POSA have been motivated to do so. Salah expressly discourages use of deep learning, and teaches away from the use of “unreliable” low contrast contours—as would be presented by an aligner in a service position. None of Maninis’s alleged advantages apply to Dr. Foroosh’s proposed combination. And, a POSA would have been discouraged from using Maninis’s COB neural network to detect generic contours because of the significant complexity and expense that would be incurred for no perceived benefit.
- Dr. Foroosh’s proposed combination does not disclose the claim limitations of claims 7 and 14. Dr. Foroosh’s proposed combination lacks the required “tooth attributes” and “tooth attribute values” in the historical images of the

learning base and in the analysis of the images assessed in the claimed methods.

- Dr Foroosh’s proposed combination does not disclose the limitations of claim 8. Dr. Foroosh proposes that the learning base comprise historical images labelled with semantic contours, which are not “image attributes” as required in Step 1’).

III. BACKGROUND

A. Education and Industry Experience

3. My qualifications are summarized below and are presented in more detail in my updated professional CV (EX2029).

4. Additionally, the following overview of my background pertains to my qualifications for providing expert testimony in this matter.

5. I have worked as a professor in the field of radiology for nine years, and have been involved with various aspects of informatics including machine learning, deep learning, and neural networks for the duration of my career.

6. I received a B.S. from Stanford University (1999) in Chemistry, graduating with honors and distinction. I then received a Ph.D. (2006) in bioinformatics and an M.D. (2008) from the University of California San Diego.

7. From 1995 to 2000, I worked part time as a programmer in the software testing development department of Autodesk, Inc. My work there led to three issued

U.S. patents on which I am named as an inventor or co-inventor. I am the lead author of *Programming Interviews Exposed*, currently in its fourth edition. This book has sold over 100,000 copies across all editions and has been translated and republished in China (Simplified Chinese), South Korea (Korean), India (English), Russia (Russian), and Poland (Polish). During my Ph.D., I created and contributed to several molecular simulation software applications, including AMBER, on which I am a named contributor for versions 8 through 10, and Interactive Essential Dynamics.

8. My post-graduate medical training included an internship in Internal Medicine at Kaiser Oakland Medical Center from 2008 to 2009, a residency in Diagnostic Radiology at University of California San Francisco from 2009 to 2013, and a fellowship in Ultrasound and Abdominal Imaging, also at University of California San Francisco from 2013 to 2014.

9. Since 2014, I have worked as a professor in the Department of Radiology and Biomedical Imaging at University of California San Francisco, initially at the assistant professor level until my promotion to associate professor in 2019. I have recently been promoted to full professor, effective July 1, 2024. In this role, I practice abdominal radiology, including reading X-rays, CT, Ultrasound, and MRI studies; and performing image-guided biopsies and intra-operative image guidance. I train residents and fellows in radiology, incorporating them into my

clinical work, as well as delivering formal lectures on a wide range of topics with a particular focus on artificial intelligence, informatics, and physics. I also conduct grant and industry-supported research, with a focus on artificial intelligence and informatics.

10. I hold two medical board certifications: Diagnostic Radiology from the American Board of Radiology and Clinical Informatics from the American Board of Preventive Medicine. In 2022, I became a Fellow of the Society of Abdominal Radiology.

11. Since 2014, I have held leadership roles in Informatics within the Department of Radiology and Biomedical Imaging, as Associate Chair of Informatics, Vice Chair of Informatics, and Associate Chair of Translational Informatics. In these roles, I have overseen the development of a shared departmental infrastructure for AI research and translational deployment, supervised an Imaging Information Technology (IT) group of approximately 20 people responsible for the clinical computer systems used for imaging, and supervised the Scientific Computing Services IT group that manages the shared research compute infrastructure for the department.

12. I hold several university-wide leadership roles in IT and artificial intelligence, including serving since 2022 as the Chair of the IT Governance Committee on Research Technology, which provides guidance for University

investment in and use of computational infrastructure for research, and serving as a co-chair of the Imaging Artificial Intelligence Oversight Committee and a member of its sister committee, the Artificial Intelligence Governance Oversight Committee, both of which review artificial intelligence prior to clinical deployment to ensure safety and efficacy.

13. I have leadership roles in artificial intelligence and informatics in national professional societies. I have been a member since 2018 and chair since 2020 of the Artificial Intelligence Committee of the Radiological Society of North America (RSNA). The RSNA is one of the largest radiology professional societies in the world, with over 54,000 members. Under my leadership, the Artificial Intelligence Committee has hosted artificial intelligence competitions on the Kaggle website every year. Several of these competitions have drawn approximately 1000 teams to compete. The committee also advises the RSNA board on AI position statements, endorsement of multi-society AI-related manuscripts, and the development of the AI Certificate education program. I have also served on the Society of Abdominal Radiology Artificial Intelligence Emerging Technology Committee since 2020.

14. Since 2018, I have been an Associate Editor and Editorial Board Member of the peer-reviewed journal *Radiology: Artificial Intelligence* (impact

factor 9.8), which focuses on applications of artificial intelligence to radiology and medical imaging.

15. I have authored over 59 peer-reviewed journal articles, 24 of which are related to artificial intelligence and machine learning, and I have given over 130 lectures, including invited plenary lectures on artificial intelligence for the German Röntgen Congress and the Colombian Congress of Radiology.

16. I have completed multiple successful research projects involving development of artificial intelligence for radiology. A highlight of my research work was development of an artificial intelligence model for detection of pneumothorax. This work was published in *Public Library of Science: Medicine* (impact factor 11.6) and licensed to General Electric (GE) for incorporation into their Critical Care Suite. I worked with GE to achieve FDA clearance for the model, and it is currently commercially available on GE X-ray machines. I am the site primary investigator for a pending National Institutes of Health R01 grant that will support development of artificial intelligence for risk stratification of colorectal cancer patients for treatment planning.

17. Finally, as noted above, I am currently named as an inventor or co-inventor on three issued U.S. patents related to computers and software.

18. For at least these reasons, and based on the additional education and experience set forth in my CV, I believe I am qualified to provide the opinions in this declaration.

19. I have offered testimony from the perspective of a POSA regarding the '248 patent in the related District Court proceeding (*Dental Monitoring SAS v. Align Technology, Inc.*, 3:22-cv-07335-WHA), and I understand that Petitioner has not objected to my qualifications.

B. Compensation

20. I am being compensated at my customary rate of \$800 per hour. My opinions are objective, and compensation for my services is not contingent on my opinions or the outcome of this proceeding.

C. Materials and Other Information Considered

21. My opinions are based on my years of education, research, and experience, as well as my investigation and study of relevant materials. In forming my opinions, I have considered the materials I identify in this declaration.

22. In particular, in forming my opinions, I have considered the Petition (Paper 2) and the declaration of Dr. Foroosh (EX1002), and the exhibits cited therein.

23. My analysis of the materials in this investigation is ongoing and I will continue to review any new material as it is provided. This declaration represents

only those opinions I have formed to date. I reserve the right to revise, supplement, and/or amend my opinions stated herein based on new information and on my continuing analysis of the materials already provided.

IV. UNDERSTANDING OF THE LAW

24. I am not an attorney and offer no opinions on the law. I have relied on instructions from counsel as to the applicable legal standards to use in arriving at my opinions in this declaration. Below is my understanding regarding the relevant legal standards based on those instructions.

A. Person Of Ordinary Skill in the Art

25. I understand that, in the context of a validity or patentability analysis, the person of ordinary skill in the art (“POSA”) is a hypothetical person who looks to the prior art at the time of the invention and without the benefit of hindsight. I understand that the POSA is presumed to have knowledge of all references that are sufficiently related to one another and to the pertinent art, and to have knowledge of all arts reasonably pertinent to the particular problem that the claimed invention addresses. I further understand that factors that may be considered in determining the level of ordinary skill in the art may include: (1) type of problems encountered in the art; (2) prior art solutions to those problems; (3) rapidity with which innovations are made; (4) sophistication of the technology; and (5) educational level

of active workers in the field. My specific understanding of a POSA with respect to the field of the '248 patent in particular is summarized below.

B. Date of Invention

26. I understand that the priority date credited to an alleged invention is relevant to an analysis of the validity or patentability of that alleged invention at least because it determines which art is prior art to the alleged invention (for my understanding of the categories of prior art see the “prior art” section below for more details).

27. I understand that when the date of invention is based on the filing of a patent application, the date of invention is the earliest effective filing date of a patent's claims. I understand that the earliest effective filing date of a patent's claims may predate the actual filing date of the U.S. application for that patent. For example, if an application is a continuation of an earlier application (a “parent application”), the earliest effective filing date of the claims resulting from the later application may be the filing date of the parent application if the parent application's disclosure satisfies the enablement and written description requirements for those claims. Similarly, if a non-provisional patent application claims priority to an earlier provisional application, the earliest effective filing date of the claims resulting from the non-provisional application may be the filing date of the provisional application

if the provisional application's disclosure satisfies the enablement and written description requirements for those claims.

C. Prior Art

28. I understand that prior art includes items that were publicly known or that were used by others before the effective filing date of the alleged invention. Prior art also includes any printed publications, published patent applications, or patents that were published or effectively filed before the effective filing date of the alleged invention. However, I understand that prior art does not include a public disclosure, publication, or patent that describes the inventor's own work if it was disclosed less than one year before the date of invention. I further understand that the prior art must be assessed from the perspective of a person of ordinary skill in the art.

D. Claim Construction

29. I understand that claim construction is a matter of law and that claim terms are interpreted as a POSA would have understood them at the time of the invention, in view of the patent and evidence related thereto.

30. I understand that there are two types of claims: independent claims and dependent claims. I understand that an independent claim stands alone and includes only the limitations it recites. I further understand that a dependent claim, on the other hand, is a claim that depends on another claim, and that dependent claims

include all of the limitations recited in the dependent claim as well as any limitations included in the claim(s) on which it depends.

31. I understand that, to determine the meaning of claim terms, one should first consider the intrinsic evidence, which includes the claim language itself, the patent specification, and the patent's prosecution and any post-issuance history before the Patent Office. For example, the patent specification may show that the inventor used words or terms in a manner inconsistent with their plain and ordinary meaning. Specifically, I understand that where the specification reveals a special definition given to a claim term by the inventor that differs from the meaning it would otherwise possess, the inventor's lexicography governs. I understand that the prosecution history of the patent may also provide guidance in construing a claim term. For example, the prosecution history may show that the patent applicant might have limited the scope of some or all of the claims during prosecution, either affirmatively or by implication.

32. I understand that if the intrinsic evidence is not conclusive regarding the meaning of a particular claim term, extrinsic evidence may also be used to determine its meaning. I understand that extrinsic evidence may be used, for example, to help determine what a POSA at the time of the invention would have understood the claim term to mean. Extrinsic evidence may include, for example, dictionaries, technical treatises, journals, articles, or expert testimony.

E. Patentability

33. I understand that on March 16, 2013, the America Invents Act (AIA) went into effect, changing American patent laws and rules as applied to patents filed on or after March 16, 2013. I understand that the application from which the '248 patent issued was filed after March 16, 2013.

34. I understand that generally the party challenging the validity or patentability of an issued U.S. patent bears the burden of proof to demonstrate that such patent is invalid or unpatentable. I understand that in this proceeding, Petitioner has the burden to show that the '248 patent claims are unpatentable by a preponderance of the evidence.

F. Obviousness

35. I understand that a claim is invalid or unpatentable for obviousness if the differences between the invention and the prior art are such that the claimed subject matter as a whole would have been obvious at the time the invention was made to a person of ordinary skill in the art to which the subject matter pertains.

36. I understand that obviousness is ultimately a question of law based on a determination of a number of factual issues. Those factual issues relate to: (1) the scope and content of the prior art, (2) the differences between the prior art and the claimed invention as a whole, (3) the level of ordinary skill in the art at the time the invention was made, and (4) secondary considerations reflecting indicia of

nonobviousness of the subject matter sought to be patented, as discussed further below.

37. I understand that the obviousness inquiry takes place at the time of the invention. Therefore, care must be taken to avoid the impermissible use of hindsight in an obviousness analysis. I understand that it is improper to use the invention as a plan or template for hindsight reconstruction of elements of the prior art to piece together the claimed invention.

38. It is my further understanding that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was known independently in the prior art. There must be an apparent reason why it would have been obvious to modify the prior art to arrive at the claimed invention with a reasonable expectation of success of doing so. I understand that to avoid impermissibly applying hindsight, one insight into the obviousness determination is whether there is a teaching, suggestion, or motivation in the prior art that would lead a POSA to combine the elements.

39. I also understand, however, that a person of ordinary skill in the art is also a person of ordinary creativity, and not an automaton. Thus, I understand that the ordinary creativity of a skilled artisan may provide a motivation to combine references. I understand that common sense may also provide a motivation to combine references. I understand that the obviousness analysis need not seek out

precise teachings directed to the specific subject matter of the challenged claim, but instead can take account of the “ordinary innovation” and experimentation that does no more than yield predictable results, which are inferences and creative steps that a person of ordinary skill in the art would employ.

40. I further understand that the patent challenger bears the burden to show that a person of ordinary skill in the relevant field had a reason and ability to combine the elements in the manner claimed and a reasonable expectation of success in doing so when asserting obviousness in view of a combination of references.

41. I further understand that where a patent claims a structure already known in the prior art that is altered by the mere substitution of one element for another known in the field, the combination must do more than yield a predictable result. Similarly, I understand that when a patent simply arranges old elements with each performing the same function it had been known to perform and yields no more than one would expect from such an arrangement, the combination is obvious. That is, the claimed improvement must be more than the predictable use of prior art elements according to their established functions.

42. I further understand that if a technique has been used to improve one device, and a POSA would recognize that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill.

43. I further understand that a reference may be said to “teach away” from the invention when a POSA, upon reading the reference, would be discouraged from following the path set out in the invention, or would be led in a direction divergent from the path that was taken by the patentee. For example, a reference teaches away if (1) a combination would produce a seemingly inoperative device, or (2) the reference leaves the impression that the product would not have the desired property. However, mere expression of a general preference for one alternative, even if over another alternative, does not teach away if it does not criticize, discredit, or otherwise discourage investigation into the claimed invention.

44. I further understand that a reference qualifies as prior art for an obviousness determination where that reference is analogous to the claimed invention. I understand that art is analogous to the claimed invention where it is either (1) from the same field of endeavor, regardless of the problem addressed or (2) even if not from the same field, is still reasonably pertinent to the particular problem with which the inventor is involved. I understand that a field is analogous to the claimed invention where a POSA would reasonably be expected to look to that field for a solution to the problem at hand. I understand that a reference is reasonably pertinent if it is one which, because of its subject matter, logically would have commended itself to an inventor’s attention in considering his problem.

45. I also understand that the following are examples of rationales that may support a conclusion of obviousness:

- Combining prior art elements according to known methods to yield predictable results;
- Simple substitution of one known element for another to obtain predictable results;
- Use of known technique to improve similar devices/products in the same way;
- Applying a known technique to a known device/product ready for improvement to yield predictable results;
- “Obvious to try”—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations are predictable to one of ordinary skill in the art;
- Some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill in the art to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention with a reasonable expectation of success.

V. THE '248 PATENT

A. The Effective Filing Date of the '248 Patent

46. The '248 patent was filed as U.S. Application No. 16/030,137 on July 9, 2018, claiming priority to French Patent Application No. 1756956, filed July 21, 2017. I understand that Petitioner has not challenged the entitlement of the '248 patent to this priority date. Pet. 12; EX1002, ¶¶26.

B. Person of Ordinary Skill in the Art

47. In my opinion, a person of ordinary skill in the relevant field as of July 21, 2017 would have had a bachelor's degree or higher in computer science, bioinformatics, or a related engineering discipline, and several years of work experience relating to the development of machine learning or artificial intelligence models or algorithms. A POSA would have also worked as part of a multi-disciplinary team, which would have included individuals familiar with medical imaging, dental imaging, dentistry, and/or orthodontics.

48. Dr. Foroosh contends that “a person of ordinary skill in the relevant field as of July 21, 2017, would typically have an advanced degree, such as a Masters' Degree or Ph.D. in computer science or a related field, with at least two years of work in computer vision, image analysis, medical imaging, or related areas. A POSA could have less education but significant professional experience in one or more of these fields. A POSA could be a member of an interdisciplinary team

including persons with backgrounds such as electrical engineering, computer science, optical imaging, and dentistry or orthodontics, and could call upon the knowledge of other team members as appropriate.” EX1002, ¶¶61.

49. In my declaration in support of Patent Owner’s Preliminary Response, I opined that these definitions are not materially different. EX2002, ¶¶47-49. While in my opinion a POSA **would** be part of a multidisciplinary team, and Dr. Foroosh’s definition provides that a POSA only **could** be part of such a team, Dr. Foroosh’s definition still recognizes that a POSA could consult with individuals having experience in dentistry and/or orthodontics **as appropriate**.²

50. As Dr. Foroosh recognizes, Kim, Invisalign, and Salah relate to dental and/or orthodontic treatment, and Kim and Salah discuss dental imaging, and in several respects these references are directed to and/or assume the knowledge of dental professionals. EX1002, ¶¶72-74; *see also infra* §§ VI.A-C, VII-VIII. Invisalign targets orthodontic and/or dental professionals directly. As Dr. Derakhshan alleges in her declaration, she “provided orthodontists and dentists” with Invisalign as “part of [her] duties at Align” to “provide[] instruction to orthodontists and dentists to certify them in the use of ... Invisalign clear plastic aligners.” EX1011, 1-2. As such, it is my opinion that it would be “appropriate” for

² All emphasis is added unless noted otherwise.

a POSA to consult with individuals having experience in dentistry and/or orthodontics to understand the teachings of Kim, Invisalign, and Salah. Accordingly, I have consulted with Dr. Budi Kusnoto, D.D.S., M.S. regarding the dental and/or orthodontic aspects of Kim, Invisalign, and Salah in forming my opinions herein, including as specifically noted throughout my report.

C. Background on Aligner Treatment

51. I have reviewed the description of aligner treatment in the '248 patent. EX1001, 4:52-62, 17:39-53, 27:26-47, 27:55-67, 28:1-6. According to the '248 patent, aligners are one type of orthodontic device designed to exert gentle pressure and thereby reposition a patient's teeth. *Id.*, 4:52-62. The '248 patent further explains that as of July 2017, aligner treatment generally began with a patient attending an in-person consultation with an orthodontist or other dental professional who would conduct scans of the patient's starting dentition and determine the desired final positioning. *Id.*, 27:26-30. Based on these initial scans, a sequence of aligners would be designed and fabricated by one of several third-party companies, such as Petitioner. *Id.*, 27:31-46. Once the patient received their aligners, the patient would wear one aligner for a predetermined interval of time and then change to the next aligner to continue shifting their teeth according to the treatment plan. *Id.*, 27:46-47. At that time, effective aligner treatment required frequent in-person orthodontic examinations to determine whether treatment was on track. *Id.*, 17:39-49, 27:55-59.

This reliance on periodic in-person examinations meant that treatment issues could go undiagnosed for multiple weeks between check-ins. *Id.*, 17:48-53, 28:1-6. When an orthodontist did identify significant issues with the patient’s progress at an in-person examination, the patient’s teeth would have to be rescanned, a new series of aligners manufactured and the old ones discarded. *Id.*, 27:60-67. As a result, the ’248 Patent states that aligner treatment was often delayed and prolonged. *Id.*, 27:66-67.

D. Introduction to the ’248 Patent

52. The ’248 patent relates to new and improved methods of “assessing the shape of an orthodontic aligner” for purposes of aligner treatment. *See, e.g.*, EX1001, 32:6-23, *see also, e.g., id.*, 28:34-38 (describing the invention as “considerably simplify[ing] the assessment of the good suitability of the aligner to the treatment, while making the assessment particularly reliable”); *see also id.*, 30:60-62 (“The suitability of the aligner can be assessed in the context of an orthodontic treatment (separation [*i.e.*, space between the aligner and the teeth] compatible or not with the orthodontic treatment)”). As explained in the ’248 patent, the new method “makes it possible, from simple photos or a simple film, to determine whether the aligner is abnormally detached” and, in some cases, even “determine the regions in which the aligner is separated from the teeth and assess the scale of this separation.” *Id.*, 31:30-37.

53. In particular, the '248 patent discloses a method comprising “acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called ‘analysis image.’” *Id.*, 28:13-15. This step can be “performed at the moment when the assessment of the shape of an aligner is desired, for example more than 1 week after the start of the treatment with the aligner.” *Id.*, 30:27-30. The analysis image can be captured remotely, for example, on a patient’s cellphone. *Id.*, 15:30-35; *see also, e.g., id.*, 19:41-44, 25:60-63.

54. The “analysis image” is then analyzed by a specific “deep learning device,” preferably a neural network, that was trained on a specific dataset, a “learning base.” *Id.*, 1:39-43; *see also, e.g., id.*, 1:66-2:1, 16:5-6, 24:3-5. The '248 patent describes “training” as a process where, “[b]y presenting, as input for the deep learning device, historical images (images+descriptions), the deep learning device progressively learns to recognize patterns on an image.” *Id.*, 16:38-41; *see also id.*, 16:46-48 (“Through its training in the step 2), the deep learning device is capable of analyzing the analysis image and of recognizing said patterns therein.”). A deep learning device “trained by means of a learning base” is thus a deep learning device that has undergone the process of training using a collection of images and corresponding attribute values for those images. *See id.* The '248 patent further

discloses a “detailed analysis method” and a “global analysis method” using a deep learning device. *E.g., id.*, 1:29-3:18 (detailed), 3:19-4:39 (global).

55. In the detailed analysis method, the “analysis image” is analyzed by the deep learning device “so as to determine a value for at least one tooth attribute of an ‘analysis tooth zone’ representing, at least partially, a tooth on said analysis image.” *Id.*, 32:14-19; *see also id.*, 16:5-45. The tooth attribute relates “to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image.” *Id.*, Abstract, 28:21-24, 32:19-21. The tooth attribute value thus “provides an information item relating to the shape of the aligner relative to the shape of the teeth of the patient.” *Id.*, 30:42-44. It “can be a measurement of the separation, for example a measurement of the maximum separation, or of the mean separation for the tooth represented by the historical tooth zone.” *Id.*, 30:45-48; *see also id.*, 16:46-17:4. Alternatively, the tooth attribute value can be, *e.g.*, a tooth number, a tooth type, and a tooth shape parameter. *Id.*, 7:64-8:7, 28:48-61, 30:42-54; *see also id.*, 16:46-17:4.

56. The deep learning device is trained by means of a learning base comprising, for example, “more than 1000 images of dental arches,” or “‘historical images,’ each historical image representing an aligner worn by a ‘historical’ patient.” *Id.*, 34:23-31; *see also, e.g., id.*, 15:65-16:4, 16:37-43. Each “historical image” comprises “one or more zones each representing a tooth, or ‘historical tooth

zones’ to each of which, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, a tooth attribute value is assigned.” *Id.*, 34:23-31; *see also, e.g., id.*, 1:33-38.

57. In one embodiment of the detailed analysis method, the deep learning device analyzes the analysis image—the image of the patient’s dental arch with an aligner in place (*i.e.*, in a service position)—to determine “at least one probability relating to: the presence, in a location of said analysis image, of an analysis tooth zone; and the attribute value of the tooth represented on that said analysis tooth zone.” *Id.*, 34:34-40. As a function of that probability, an amplitude of the separation between the patient’s teeth and the aligner is determined. *Id.*, 34:41-42.

58. The analysis may also be performed globally, without identification of the individual situation of each of the teeth represented. *E.g., id.*, 3:19-4:39; *see also id.*, 17:9-18:53. In this global analysis method, the analysis image is analyzed by a specially trained “deep learning device” so as to determine a value for an image attribute relating to a separation between at least one tooth and the aligner. *E.g., id.*, Abstract, 29:31-53. The image attribute relates to the image as a whole and includes the maximal separation, the mean separation, and the overall acceptability of the separation between the teeth and the aligner. *Id.*, 28:48-61; *see also id.*, 31:20-29 (“For example, the value of the image attribute may be ‘globally acceptable’ or

‘globally unacceptable’. The value of the image attribute can also, for example, be a measurement of the separation, for example a measurement of the maximum separation, or of the mean separation between the teeth represented on the analysis image and the aligner.”).

59. As discussed above, in the embodiment according to the **detailed** analysis method, the learning base comprises historical images of dental arches representing an aligner on a historical patient “comprising one or more zones each representing a tooth, or ‘historical tooth zones’, to each of which is assigned, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, **a tooth attribute value.**” *Id.*, 28:62-29:13. And in this embodiment, the method for monitoring an orthodontic aligner comprises “submission of the analysis image to the deep learning device for the deep learning device to determine at least one probability relating to the presence, at a location in said analysis image, of an analysis tooth zone, and the **tooth attribute value** of the tooth represented on said analysis tooth zone,” and “determination, as a function of said probability, of the presence of a separation between the aligner and the tooth represented by said analysis tooth zone, and/or of an amplitude of said separation.” *Id.*, 29:16-26; *see also id.*, 15:57-16:6, 16:37-54, 30:35-53. In other words, the learning base is trained on images labeled with tooth zones and tooth attribute values, and it is used in an analysis to determine tooth zones

and tooth attribute values. By contrast, in the embodiment according to the *global* analysis method, the learning base is trained on images labeled with *image attribute values*, and it is used in an analysis to determine *image attribute values*. *Id.*, 17:9-34; *see also id.*, 30:13-29.

60. Preferably, an “assessment of the suitability of the aligner” is made as a function of the tooth or image attribute value. *E.g., id.*, Fig. 9, 28:29-31, 32:26-27, 34:60-61. For example, a search may be “conducted to see if the separation of the aligner with at least one tooth exceeds an acceptability threshold” and, if so, the patient’s aligner is replaced by a better-suited one. *Id.*, 30:56-59. As a function of the suitability assessment, an information message may be sent to the patient and/or the orthodontist. *Id.*, 32:35-36; *see also, e.g., id.*, 26:35-27:3. “The orthodontist can then use this information, possibly in combination with additional information, for example the age of the patient or the duration for which the aligner has been worn, to establish a diagnosis and, if necessary, decide on a suitable treatment.” *Id.*, 31:7-11.

E. The State of the Art Relating to Neural Networks

61. In the context of image recognition, as of July 2017, it was known that neural networks were algorithms that had been trained to evaluate certain images and output probabilities associated with attributes they had been trained to recognize in an image. A POSA would have understood that neural networks are not general-

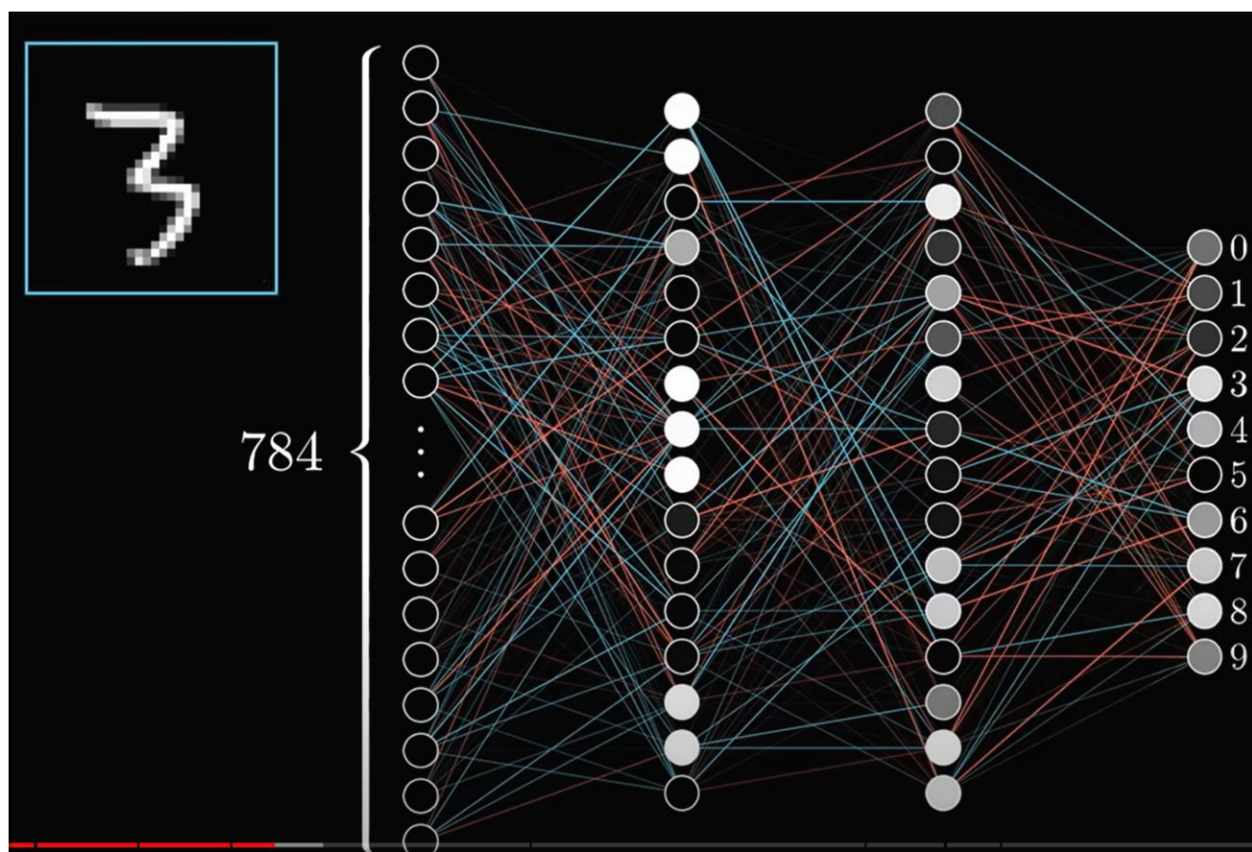
purpose tools: rather, they are only useful once they have been trained, and can only recognize what they have been trained to recognize. In other words, it was known that each neural network trained for an image recognition task is unique and specific to the annotated images on which it has been trained, and its success (then and now) depends on the quality of that training and the challenge(s) of the attributes it is tasked with recognizing.

62. As a simple and classic example, a neural network designed to recognize handwritten numbers would be trained on a set of such images that had been annotated with the number they actually represent (called the “ground truth”). For example, the following images would be labeled as depicting the number “3”:



Source: <https://www.youtube.com/watch?v=aircAruvnKk&t=16s>, 0:16-0:21(last accessed July 11, 2024).

63. At the start of training, the neural network comprises a set of calculations using parameters called “weights” and “biases” that are initially set to random values and applied in several rounds of calculation or “layers,” with each influencing the next. The neural network might, for example, take the value of the brightness of each pixel in the image as an input and apply a series of calculations to determine a probability that the represented image is one of the numbers between “0” and “9,” as represented in the below graphic as a brightness associated with the potential output options:

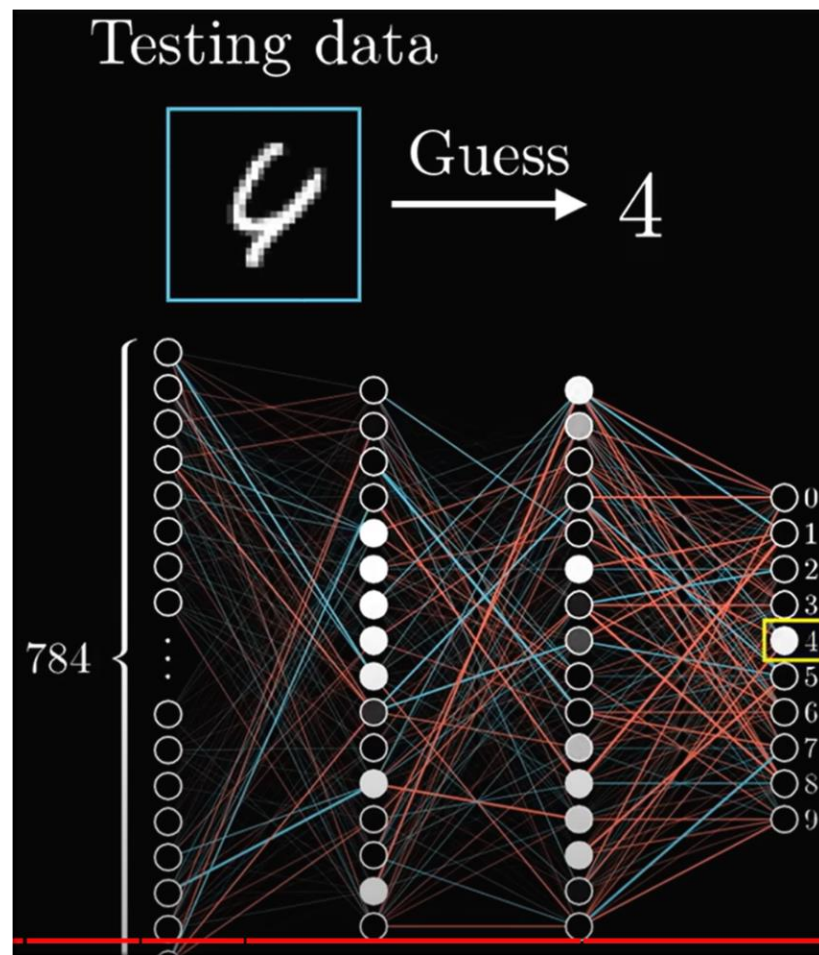


Source: https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=2, 3:34 (last accessed July 11, 2024).

64. Because the “weights” and “biases” begin at a random starting point, the probabilities are initially not a good predictor of the number depicted in the image. This is depicted above by the neural network assigning similar probabilities to the possibility that the handwritten “3” is a “3,” a “4,” or one of the numbers “6” through “9.”

65. However, by mathematically adjusting the weights and biases after each training example such that it will make a better prediction for that individual

example the next time (*i.e.*, training by “backpropagation”), the accuracy of its predictions is slowly improved. Through thousands of examples and corrections, or millions of examples in the case of more complex applications, the neural network can potentially learn to output a usefully accurate probability associated with the attribute on which it was trained (here, what number is handwritten in the image), as illustrated below:



Source: https://www.youtube.com/watch?v=IHZwWFHwa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=2, 2:34 (last accessed July 11, 2024).

66. In the simple example above, while the input to the neural network is a vector representing a bitmap (*i.e.*, a series of values representing the brightness/coloration of each pixel in an image), the output is a probability associated with a single value for the entire image (*i.e.*, the number depicted in the image). In the context of the '248 patent, this example represents the “global analysis” approach, and the number depicted in the image would be an example of an “image attribute.” As discussed below in the context of Maninis, a neural network might instead output a vector representing a bitmap—for example, a bitmap of the same size as the input image, with values for each pixel representing a probability that each pixel is part of a contour or segmentation³ associated with an object.

³ As of 2017, a POSA would have understood that “boundaries” or “contours” are the outlines of objects depicted in an image. A POSA would have known that contour (or edge) detection is the process of detecting the location of those contours (or edges) in an image. In practical terms, this means identifying whether each pixel in an image is part of a contour or not. A segment is the region within a contour. Similarly, a POSA would have known that “semantic segmentation” is the labeling of pixels in an image as associated with a particular object and “semantic boundaries” are the contours (or edges) that define a particular identified object.

67. A neural network must be structured to have the appropriate output before it is trained and cannot be trained on data that is the wrong size/shape for the network. For example, in the above example, the output vector is only 10 values long—one for each numeral between 0 and 9. If the output was instead a bitmap representing a contour or segmentation, the output vector would be 784 values long—one for each pixel in the image. Because the neural network must be trained by comparing its output to the “ground truth” data in the training image (so its weights and biases can be adjusted going forward), a neural network can only be trained on images having annotations that match the structure of the neural network—the above examples are not interchangeable without re-designing the structure of the neural network.

VI. OVERVIEW OF THE ASSERTED ART

A. Kim

68. U.S. Patent Publication No. 2017/0281313 to Kim is entitled “Method and System for Providing Remote Teeth Alignment Services and Computer-readable Medium on which the Method is Recorded.” EX1005,[Kim]. Kim was filed on April 5, 2016, and was published after the ’248 patent priority date, on October 5, 2017. *Id.*

69. As Kim explains, as of 2016, “[u]sing the [then] conventional clear-aligner treatment method, an orthodontic patient should regularly visit a dental clinic

for a long period of time to allow a dentist to accurately check the state of his/her teeth and determine a currently appropriate step of a clear aligner.” *Id.*, [0008]. Thus, Kim is directed to the problems that could occur when a patient was not capable of frequently visiting the orthodontist by providing a way to assess “teeth alignment” remotely. *Id.*, [0009], [0010].

70. Kim thus relates to a method for remotely determining the alignment of teeth by “providing remote teeth alignment services” to assess whether treatment is progressing correctly, and not for evaluating aligner fit. *Id.*, [0010] (disclosing “a method and system for providing remote teeth alignment services”). That method involves an orthodontic patient sending the orthodontist “teeth alignment information using a device, e.g., a smartphone.” *Id.* Kim teaches only photographs taken by the patient, using the “orthodontic patient terminal.” *E.g., id.*, [0013], [0041], [0056], [0058], Fig. 6. Kim is also clear that the orthodontic patient terminal is a personal electronic device of the patient, such as “a personal computer, a mobile device, a personal digital assistant (PDA), a smart pad, a laptop computer, a smart camera, a smart camcorder, an electronic book, a smart scanner, or the like.” *Id.*, [0042]. Thus, Kim does not disclose that photographs should be taken by a trained professional, using correct lighting and camera equipment. Kim asserts that, on the basis of the teeth alignment information, the orthodontist may “remotely and accurately diagnose the condition of the patient based on the teeth alignment

information, and may remotely provide information about an accurate step of a clear aligner to be currently worn by the orthodontic patient.” *Id.*, [0010].

71. Specifically, “[t]he teeth alignment report information” sent by the patient may include “a front-view teeth photo, a bottom-view low-angle teeth photo, a front-view automatically-recognized mouth gag photo, a front-view automatically-recognized face contour photo, a left-side-angle automatically-recognized mouth gag photo, a left-side-angle automatically-recognized face contour photo, a right-side-angle automatically-recognized mouth gag photo, a right-side-angle automatically-recognized face contour photo, a collection of top, bottom, left, and right side photos surrounding a central front photo, a low-angle upper jaw photo, a high-angle lower jaw photo, and combinations thereof, which are taken using a camera of the orthodontic patient terminal.” *Id.*, [0013]; *see also id.*, [0056] (discussing acquisition of images of the patient’s teeth, not the aligner on the patient’s teeth), [0081] (same), [0087] (same), Figs. 8-13 (showing only images of the patient’s teeth without aligners), 16 (same), 19 (same), 23 (same). These are all images of the patient’s teeth taken without an aligner in a service position. While Kim discloses the use of programs to “display[] a virtual face contour line on a screen” to guide camera positioning and automatically capture photographs, it does not disclose doing so while the patient wears their aligner in a service position. *Id.*, [0014].

72. I understand from Dr. Kusnoto that, as of 2017, assessing the fit of an aligner and assessing the alignment of teeth were two different analyses. At that time, assessing the fit of an aligner involved checking that the aligner provided adequate and effective pressure on the patient's teeth to bring about the desired movement of the teeth. During an in-person appointment, this typically involved a combination of visual and manual inspections that were performed by the dental professional. By contrast, assessing the alignment of the patient's teeth in 2017 involved an overall evaluation of the patient's dentition to determine whether the orthodontic treatment was progressing as planned. Such an assessment of teeth alignment was based on tooth positions, and it was a standard part of orthodontic treatment in 2017 that was not specific to the type of treatment (braces, aligners, retainers, etc.). I further understand from Dr. Kusnoto that clear aligners can mask the teeth, which would make it difficult to analyze teeth positions from an image in which an aligner is on the patient's teeth. Thus, a POSA would not have been motivated to use an aligner image to determine **teeth alignment**.

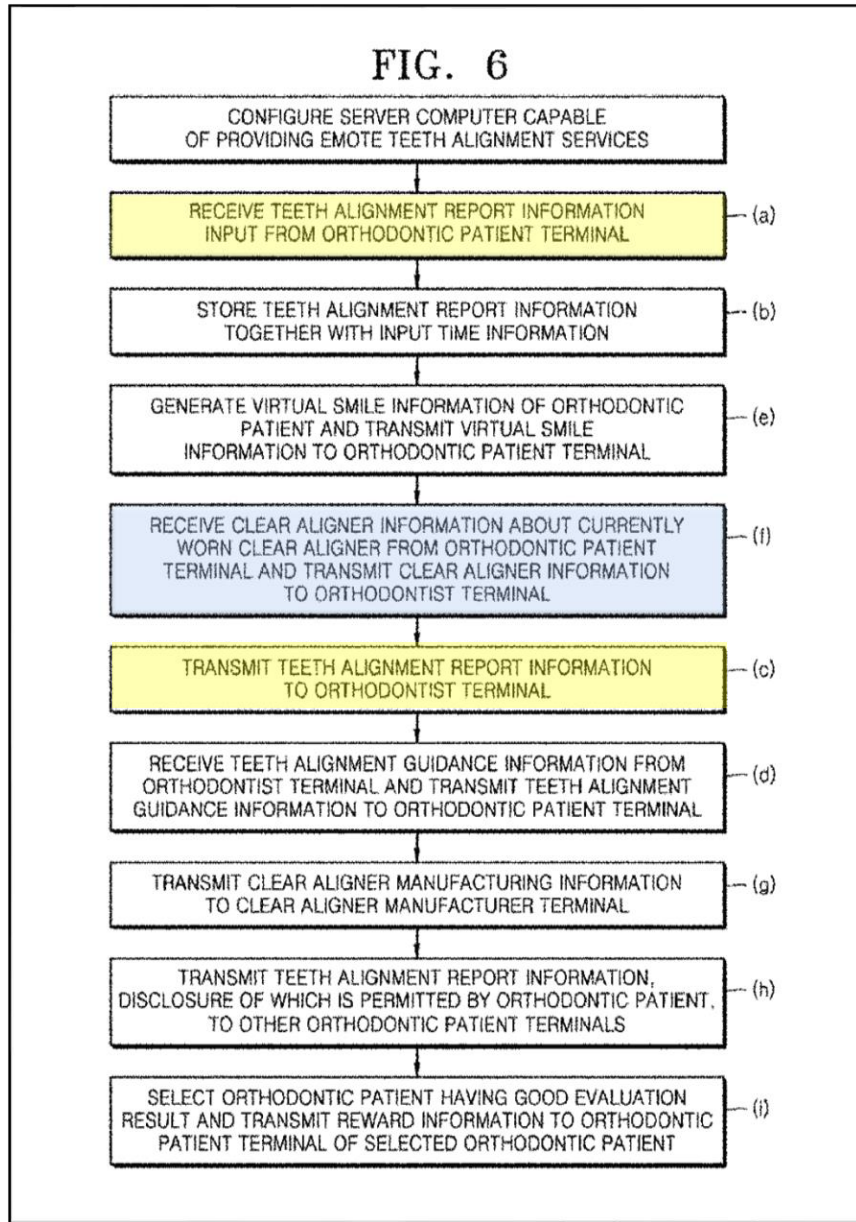
73. Dr. Foroosh nonetheless opines that "Kim teaches acquiring an analysis image of a patient's teeth wearing an aligner" to evaluate aligner fit. EX1002, *e.g.*, ¶¶84, 87. In support of this opinion, he relies on Paragraph [0066] of Kim, which states as follows:

The clear aligner information transmission program 69 may be a program for receiving the clear aligner information about a currently

worn clear aligner from the orthodontic patient terminal 10, storing the clear aligner information, and transmitting the clear aligner information to the orthodontist terminal 20 in such a manner that the orthodontist checks a photo or a video of the clear aligner currently worn by the orthodontic patient, and determines whether the clear aligner is normal and corresponds to an appropriate step.

EX1005, [0066]. For the reasons discussed below, I disagree with Dr. Foroosh, and it is my opinion that Kim does not relate to a method for assessing the fit of an aligner and it does not disclose “acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient,” as required by all challenged claims.

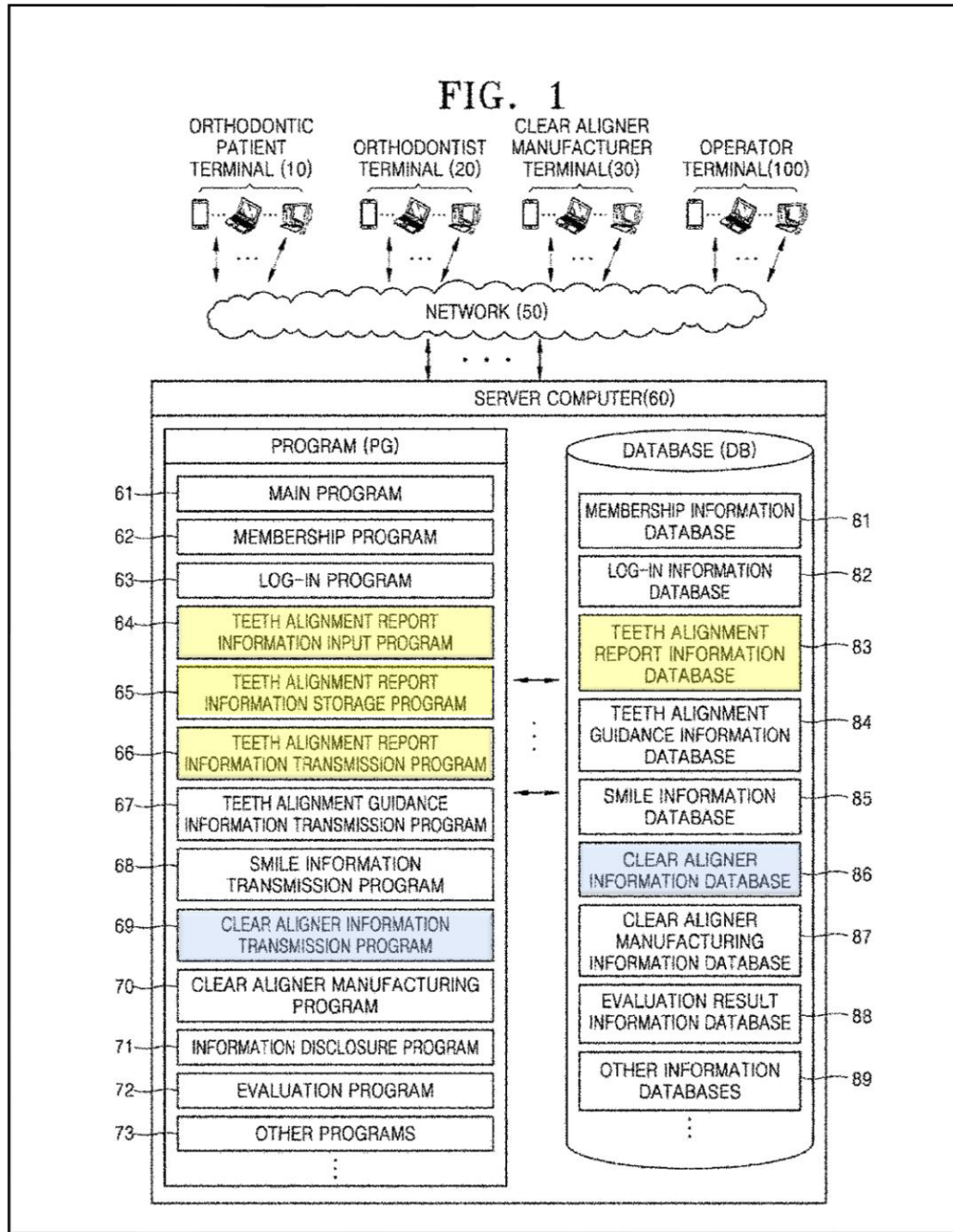
74. Kim draws a clear distinction between the “clear aligner information” referred to Paragraph [0066] and “teeth alignment information.” For example, in Figure 6 of Kim (shown below), in different steps of the methods of Kim, the server “receive[s] teeth alignment report information input from orthodontic patient terminal” and separately “receive[s] clear aligner information about currently worn clear aligner from orthodontic patient terminal.” EX1005, Fig. 6 (steps (a) and (f), highlighted below). Likewise, Kim discloses that “teeth alignment report information” and “clear aligner information” are transmitted separately to the orthodontist terminal. *Id.*, Fig. 6 (steps (c) and (f), highlighted below).



75. Kim further explains that the teeth alignment information is collected by patients and transmitted to orthodontists using a “teeth alignment report information input program” and “teeth alignment report information transmission program.” *Id.*, [0055], [0060], Fig. 2 (illustrating teeth alignment report information input program), Fig. 3 (illustrating teeth alignment report information transmission

program). That teeth alignment information is stored in a “teeth alignment report information database.” *Id.*, [0073]. By contrast, the *clear aligner information* is maintained in a “clear aligner information database” and transmitted to orthodontists “using the clear aligner information transmission program.” *Id.*, [0073]-[0074].

76. The fact that teeth alignment information and clear aligner information is collected, transmitted, and stored separately is plain from Figure 1 of Kim (shown with highlighting below):



77. As explained above, the “teeth alignment report information” sent by the patient may include photographs of the patient’s teeth. *Id.*, [0013]; *see also id.*, [0056] (discussing acquisition of images of the patient’s teeth, not the aligner on the patient’s teeth), [0081] (same), [0087] (same), Figs. 8-13 (showing only images of

the patient's teeth without aligners), 16 (same), 19 (same), 23 (same). However, Kim never suggests that the "teeth alignment report" includes information about the aligner, let alone a photograph of the aligner in a service position. *Id.*, [0013], [0056], [0081], [0087]. The sample illustrations of the teeth alignment report likewise do not show the aligners. *E.g., id.*, Figs. 8-13, 19; *id.*, [0083] (describing Figure 13 as the teeth alignment report information checked and transmitted by the patient), [0085] (describing Figure 19 as the teeth alignment report information transmitted to the orthodontist).

78. According to Kim, the teeth alignment information, not the clear aligner information, is used by the orthodontist to diagnose patients and provide aligner advice based on the position of the teeth, not the aligner fit itself. *Id.*, [0010] ("The present invention provides a method and system for providing remote teeth alignment services, by which an orthodontist may ... remotely and accurately diagnose the condition of the patient **based on the teeth alignment information**"); *see also id.*, [0085] ("the orthodontist may check the transmitted **teeth alignment report** information per patient using the orthodontist terminal"), [0087] ("the orthodontist may remotely provide accurate diagnosis by checking [the various photos] capable of showing the **teeth alignment state** of the orthodontic patient"); *see also id.*, [0063] ("the orthodontist accurately determines a **current teeth alignment state** ... and advises, for example, an appropriate step of a clear aligner to

be currently worn by the orthodontic patient.”). In fact, Kim discloses that, according to its “embodiments of the present invention, an orthodontist may obtain accurate teeth alignment information using a device, e.g., a smartphone, of an orthodontic patient, may remotely and accurately diagnose the condition of the patient ***based on the teeth alignment information.***” *Id.*, [0093]. There is no suggestion in Kim to use the clear aligner information for diagnosis or to monitor tooth positions.

79. Kim similarly discloses that the “clear aligner manufacturing program” may transmit clear aligner manufacturing information to the clear aligner manufacturer terminal “upon determining that an additional or modified clear aligner is needed, based on the ***teeth alignment report information.***” *Id.*, [0067]; *see also id.*, [0020] (same), [0090] (same). I understand that, at deposition, Dr. Foroosh agreed that Kim discloses diagnosis, treatment, and manufacturing based on the teeth alignment information. EX2054, 13:11-19 (“**Q.** And Kim generally relates to diagnosing and treating patients based on those patient-created photographs in what Kim calls a ‘teeth alignment report’; right? **A.** Yes. **Q.** And Kim also discloses that new aligners can be manufactured based on that teeth alignment report? **A.** Yes.”).

80. I understand from Dr. Kusnoto that an orthodontist would have thus understood Kim to instruct orthodontists to look at the teeth alignment information report—containing photos of the patients’ teeth without an aligner—for purposes of

assessing how the treatment is progressing. I further understand from Dr. Kusnoto that aligners can make it difficult to see the exact tooth positions. Thus, a POSA would not have been motivated to use aligner images to assess teeth alignment. I further understand that, in 2017, the position of the teeth (*e.g.*, as captured in images or models) was used to design and manufacture aligners, and that the progress of treatment was evaluated by assessing tooth alignment. Thus, I understand that using images of teeth alone, as in Kim, would have been consistent with these purposes.

81. By contrast, the clear aligner information received by the orthodontist can include information about “a **currently worn** clear aligner” (EX1005, [0018]) and the orthodontist can transmit clear aligner information about “a clear aligner **to be worn** in a current step among a clear aligner set” (*Id.*, [0016]). Kim discloses that the former may, but need not, include “a photo or a video of the clear aligner currently worn.” *Id.*, [0066]. A POSA would have understood this language to refer to a photo or video of the aligner itself. That is because, in Kim, collecting and transmitting “clear aligner information” about the aligner itself is entirely separate from collecting and transmitting information about the patient’s teeth. The clear aligner information of Kim is also not used for diagnosis, but instead to “determine[] whether the clear aligner is normal and corresponds to an appropriate step,” that is, whether the aligner appears to be in working condition and the patient is using the right aligner for their current step of treatment. *Id.*, [0066].

82. To the extent Dr. Foroosh suggests that the phrase “currently worn” aligner suggests that the aligner is being worn at the time the patient takes a photograph, a POSA would not have understood the “currently worn” aligner to be worn at the exact moment of the photograph, but instead during the patient’s current phase of treatment. Kim uses the phrase “currently worn” to refer to the aligner that the patient is wearing or should be wearing during the current step in their treatment, and not when taking photographs. *E.g., id.*, [0041] (“the orthodontic patient terminal 10 may be a device of an orthodontic patient who desires to remotely receive information ***about an accurate step of a clear aligner to be currently worn*** by the orthodontic patient”); [0063] (“The teeth alignment guidance information transmission program 67 may be a program for receiving the teeth alignment report information and then transmitting the teeth alignment guidance information received from the orthodontist terminal 20, to the orthodontic patient terminal 10 in such a manner that the orthodontist accurately determines a current teeth alignment state of the orthodontist ***and advises, for example, an appropriate step of a clear aligner to be currently worn*** by the orthodontic patient.”); [0093] (“an orthodontist may obtain accurate teeth alignment information using a device, e.g., a smartphone, of an orthodontic patient, may remotely and accurately diagnose the condition of the patient based on the teeth alignment information, and ***may remotely provide information about an accurate step of a clear aligner to be currently worn by the***”

orthodontic patient, thereby greatly reducing a diagnosis time and cost of the orthodontic patient and the orthodontist”); *see also id.*, [0051] (describing “a clear aligner information transmission program 69 for receiving a clear aligner information about a currently worn clear aligner from the orthodontic patient terminal 10, storing the clear aligner information, and transmitting the clear aligner information to the orthodontist terminal 20”).

83. Kim never claims to provide a novel way to assess aligner fit from a photograph, nor explains how that could be done. A POSA further would have understood that images captured by a patient using a cellphone, according to Kim, were unlikely to be of high quality. For example, holding the phone “close to the teeth” to get a “high level of detail,” could result in “blurry” images and “optical distortion.” EX2064, [0004]. But holding the phone “too far” from the teeth, resulted in a different problem—photos that “cannot meet the orthodontic standards.” EX2064, [0004]. This is significant because I understand from Dr. Kusnoto that aligner fit issues can be difficult to detect, especially when they are just emerging. Although Kim discloses a program to provide guidance to the patient on positioning their head or teeth (*id.*, [0014]), there is no disclosure of any guidance on how to acquire an image with an aligner seated on teeth. This confirms that Kim is not directed to assessing aligner fit from an image of a patient wearing their aligner.

84. Kim also does not suggest determining a value for a tooth or image attribute relating to the separation between the teeth and the aligner, and it does not suggest any non-human analysis of the images. *See, e.g., id.*, [0085] (“Then, as illustrated in FIG. 19, the orthodontist may check the transmitted teeth alignment report information per patient using the orthodontist terminal 20.”).

85. Dr. Foroosh states that “Kim also teaches that automatic recognition of contours of a patient’s face and teeth may be employed as part of the imaging process.” EX1002, ¶69 (citing EX1005, [0014] (“As one example, Kim teaches providing a ‘teeth contour automatic-recognition photographing program for displaying a virtual teeth line on the screen of the orthodontic patient terminal and automatically recognizing and taking a photo of teeth of the orthodontic patient shown on the screen.’”)). This disclosure relates to the acquisition of images, not to the analysis of the acquired images. Kim does not suggest the use of contour detection in the analysis of any images of the patient’s dental arch. Kim also does not suggest determining a value for a tooth or image attribute relating to the separation between the teeth and the aligner, and it does not suggest any non-human analysis of orthodontic treatment from the acquired images. *See, e.g.,* EX1005, [0085] (“Then, as illustrated in FIG. 19, the orthodontist may check the transmitted teeth alignment report information per patient using the orthodontist terminal 20.”).

86. Finally, Kim does not disclose analysis of any image by a deep learning device trained by means of a learning base. In addition, for the reasons discussed above, I disagree that Kim discloses acquiring images of an aligner in a service position, and thus I also disagree with Dr. Foroosh (*e.g.*, EX1002, ¶¶102, 123, 139) that images “taken in the context of Kim” would have been appropriate to train a deep learning device to perform the image analysis claimed in the ’248 patent.

B. Invisalign

87. I have reviewed the Invisalign Clinical Monitoring Guide (“Invisalign,” EX1006). I understand that Invisalign was published by Petitioner “to help the Invisalign practitioner successfully monitor Invisalign treatments and address issues that may arise between receipt of Aligners from [Petitioner] to the completion of treatment.” EX1006, i. Invisalign teaches that, “[t]o prevent issues from occurring,” dental professionals should “perform” certain steps “***at every patient appointment.***” *Id.*, ii. For example, Invisalign discloses that dental professionals should have the “***patient[] arrive*** with the previous stage Aligner and the current Aligner” so the dental professional can evaluate any treatment concerns that may arise. *Id.*

88. Unlike the challenged claims of the ’248 patent, Invisalign discloses only a method for assessing aligner fit that is performed in person, during in-office orthodontist check-ups. Invisalign does not suggest that aligner treatment progress can be remotely assessed from an image of a patient wearing their aligner in a service

position. In fact, I understand that, until 2020, Petitioner did not offer any product for remotely assessing aligner treatment progress. EX2008.

89. I further understand from Dr. Kusnoto that a POSA would have understood Invisalign to instruct orthodontists to perform certain steps at each appointment that *could not* be performed remotely. These include, for example, Invisalign’s instruction that: “At each appointment ... [e]valuate for tight contacts with unwaxed floss and relieve with finishing strips if present.” EX1006, ii. Invisalign also instructs orthodontists to use a scaler “to trim attachment flash as well as check retention.” *Id.*, 1. I understand from Dr. Kusnoto that manipulating and evaluating aligner fit using a scaler cannot be performed virtually.

90. Invisalign includes only exemplary images of what a dental practitioner might see during an in-person appointment when the aligner does not fit properly. EX1006, 1. However, I understand from Dr. Kusnoto that the specific image from Invisalign on which Dr. Foroosh relies (Figure “A,” reproduced below) is an extreme example of an unseated aligner that is not engaged with the patient’s teeth. I further understand from Dr. Kusnoto that Figure A does not accurately reflect the spaces between the aligner and the teeth that would typically have been observed during aligner treatment in 2017 and that the spaces were generally much less perceptible than shown in Figure A below. According to Dr. Kusnoto, Invisalign’s image of “Aligner seating well” (Figure “B”, reproduced below) is a far better representation

of what orthodontists had to evaluate during the normal course of aligner treatment.

EX1006, 1.



91. Additionally, Invisalign makes no suggestion that any image was taken using a cellphone, nor that images should or could be acquired using a cellphone to assess aligner treatment. A POSA would have understood that the images in Invisalign were taken with a more sophisticated camera, such as a DSLR, and in-office with bright and controlled lighting, as I understand Dr. Foroosh has also acknowledged. EX2054, 134:10-135:24 (When asked how he “know[s] that the photograph in Invisalign is a digital photograph?,” responding “I just assume that in modern days, even in dentist office I’ve been, they are using SLR cameras. Those are all digital.”; “I was referring to SLR type of cameras that, you know, dentists use

in offices as opposed to, let's say, camera on our cell phones. I've seen that they're all digital cameras."").

92. I also understand from Dr. Kusnoto that specialized training and equipment, *e.g.*, cheek retractors, dental mirrors, and a DSLR camera, were required to acquire high quality dental images and that in 2017, it was rare for dental practitioners to use cellphones to acquire images of their patients. EX2008,[Align May 07, 2020 Press Release] (Petitioner's CEO, Joe Hogan, stating that "[e]arly on [in the 2020 COVID-19 pandemic], many doctors **began** using video calls, texts, and patient-submitted photos through a variety of platforms to help monitor patient progress, reduce in-office appointments, and ensure continuity of patient care during treatment.""). A POSA would have expected that images acquired on a cellphone would be lower resolution and lower quality than a DSLR camera, especially if they were acquired by a non-trained professional as envisaged by Kim and Salah.

93. Invisalign does indicate that a separation between the teeth and the aligner of greater than 1 mm reflects a general aligner fit issue, whereas a discrepancy less than 1 mm is a "natural deflection that should occur so the Aligner can move the tooth." EX1006, 1. But Invisalign does not disclose any method to assess the separation, other than a visual in-person approximation, and it does not provide any method and/or tool for measuring a separation, even in person. A POSA would not have been able to make any measurement from the images in Invisalign

without the ability to determine the scale of the image. This is significant because, for example, if a patient held their phone six inches from their mouth, the gap would appear to be a different size than if they held their phone a foot from their mouth. Thus, Invisalign suggests that, in the case of a separation that appears to be less than 1 mm, the dental professional should simply continue to monitor aligner fit. *See id.*

94. Invisalign discusses only the collection of images for the purposes of aligner design and manufacture, *i.e.*, before the aligners exist. Contrary to Dr. Foroosh's suggestion, there is nothing in Invisalign that indicated that such images should be taken with a cellphone. *See* EX1002, ¶102. As discussed above, the '248 patent discloses that conventional aligner treatment generally began with a patient attending an in-person consultation with an orthodontist or other dental professional who would conduct scans (or collect an impression and/or photos) of the patient's starting dentition, which were used to design and fabricate the clear aligners. *See* Section V.C. As these images were taken before the aligners were manufactured, they could not have included images of a patient wearing an aligner in a service position.

95. For the above reasons, I disagree with Dr. Foroosh that Invisalign describes acquiring "photographs of the teeth of the patient wearing the aligners." EX1002, ¶48. Unlike the '248 patent challenged claims, Invisalign does not suggest

acquisition of images showing the aligner in a service position for the purposes of assessing aligner treatment.

96. Finally, Invisalign does not disclose analysis of any image by means of a deep learning device trained by means of a learning base, including so as to determine a value for a tooth or image attribute relating to the separation between the teeth and the aligner.

C. Salah

97. PCT Publication No. WO2016066651 to Salah *et al.* (“Salah”) is entitled “Method for Monitoring Dentition.” EX1008. While Salah names five inventors (*id.*), four of whom match the inventors listed on the ’248 patent, as discussed further below, it is my opinion that Salah is directed to an invention that is different from the ’248 patent.

98. Salah teaches a method for monitoring the positioning, shape, and/or appearance of a patient’s teeth using a three-dimensional (“3D”) digital reference model, taking an updated two-dimensional (“2D”) image of the patient’s teeth, and comparing the updated image to the reference model. *Id.*, 1:1-2:19; *see also* EX2002, ¶¶88-94. In other words, Salah is directed to assessing teeth alignment, which I understand from Dr. Kusnoto is distinct from assessing aligner fit or the shape of an aligner in a service position. I understand that Dr. Foroosh similarly confirmed that “Salah is not talking about clear aligners; so obviously could not talk

about how it fits -- how a clear aligner fits or seating on the teeth.” EX2054, 117:1-4.

99. In summary, the invention in Salah “provides a method for monitoring the position and/or shape of a patient’s teeth” comprising the following steps:

- a) creation of a 3D digital reference model of at least part of an arch called an “initial reference model” and, preferably, using the initial reference, defining a 3D digital reference model of each tooth called the “tooth model”;
- b) acquisition of at least one 2D image of the patient’s arches called an “updated image”, under actual acquisition conditions;
- c) analysis of each updated image and production of an updated map relating to discriminant information;
- d) optionally, determination of gross virtual acquisition conditions approximating the real acquisition conditions for each updated image;
- e) searching for a final reference model corresponding to the positioning and/or shape of the teeth at the time of acquisition of the updated image, preferably using a metaheuristic method, and
- f) for each tooth model, comparison of the positions or shapes of the tooth model in the initial reference model and in the final reference

model so as to determine the deformation and/or displacement of teeth between the steps a) and b).

EX1008, 1:25-2:19.

100. Salah states that its invention provides a method to “accurately assess tooth displacement and/or deformation since the initial reference model was produced” based on “a simple image of the teeth, taken without precise pre-positioning of the teeth in relation to the image acquisition apparatus, such as a photograph taken by the patient.” *Id.*, 2:21-26; *see also id.*, 31:4-7 (“A kit according to the invention advantageously enables image acquisition without the need for a specialist person, in particular an orthodontist. In particular, image acquisition can be carried out by the patient himself or by one of their relatives, using a simple cellphone, anywhere, and especially outside a medical, dental or orthodontic practice.”). I have discussed each of the steps in Salah in further detail below.

1. Step a) – formation of a 3D model

101. First, in what Salah refers to as “step a),” a detailed 3D model of the patient’s teeth is formed using physical modeling (*e.g.*, “plaster model”) or “preferably ... using a professional device, such as a 3D scanner, preferably implemented by a healthcare professional, such as an orthodontist or orthodontic laboratory.” *Id.*, 19:25-20:21. This 3D model is processed to determine the orientation of the arches in relation to each other and to create models of individual

teeth (“tooth model”). *Id.*, 20:22-23:27, Figs. 3-4. These models can be stored “in a centralized database” but are “preferably” transferred to “the personal device” that will be used for image acquisition such as “the patient’s cell phone.” *Id.*, 23:9-19.

102. Salah’s accuracy depends on starting with this detailed 3D model, as it becomes the reference for further analyses of the patient’s teeth over time. By having a detailed reference model, Salah provides a point of comparison for later-taken, lower-quality images. Salah does not disclose that low quality images are useful in the absence of a detailed reference model. In other words, Salah teaches that a detailed and high-quality 3D model is a necessary starting point, and does not disclose any means of starting with patient-created photographs to achieve detailed measurements about their teeth from those photographs alone.

2. Step b) – acquisition of an “updated image”

103. Next, in what Salah refers to as “step b),” an “updated image” of the patient’s dental arch is acquired. *Id.*, 23:24-33:8. In the first section, titled “Acquisition,” Salah teaches that the “updated image” is acquired at some time interval after the model was created in step a), including preferably at times that are determined by an orthodontist or that are likely to correspond to a potentially significant change in tooth position. *Id.*, 23:28-24:20.

104. The next section titled “Image acquisition apparatus” describes the equipment used to acquire the “updated image.” *Id.*, 25:13-27:13. Salah teaches

that “[p]referably, the image acquisition apparatus is a personal device commonly available on the market, for example a cell phone” and that the “updated image” is “preferably created by the patient himself or by one of his relatives.” *Id.*, 25:14-19. I note that a cellphone is the only image acquisition apparatus disclosed in Salah that Dr. Foroosh relies on for his combination. EX1002, ¶¶52, 103. Salah also indicates that it is “preferabl[e]” that “the image acquisition apparatus includes a specialized application for implementing step b)” and that “a dental retractor is used ... as shown in Figures 5a and 5c.” EX1008, 26:19-27:4, Figs. 5a, 5c.

3. Step c) – analysis of the “updated image”

105. Next, in what Salah refers to as “step c),” each “updated image is analyzed to produce ... an updated map relating to at least one piece of discriminant information,” which Salah defines as “characteristic information that can be extracted from an image (*image feature*)” such as “contour information” or “[b]rightness information.” *Id.*, 33:9-10, 18:11-17. As a preliminary step the “updated image” may be cropped to focus on the patient’s dental arch, as shown in Figures 5b and 5d. *Id.*, 33:14-34:2, Figs. 5a-d. This cropping “is facilitated by registration marks on the updated image,” which Salah describes as taking the form of “asymmetrical shape[s],” colors that stand out against the color of the retractor, and/or shapes and colors that make them individually identifiable. *Id.*, 33:16-26. It is the registration marks, and not any image analysis to identify the patient’s dental

arch, that makes “it possible to identify the area of the updated image containing ... the teeth and gums.” *Id.*, 33:27-34:1.

106. From the updated image, an updated map presenting discriminant information is then obtained. *Id.*, 34:5-36:16. Salah indicates that “[d]iscriminant information is preferably chosen from the group consisting of contour information, color information, density information, a distance information, brightness information, saturation information, reflection information and combinations of these.” *Id.*, 34:8-10. An example of an “updated tooth contour map” is depicted in Figure 6b. *Id.*, 34:7, Fig. 6b. Salah does not disclose the use of semantic segmentation or object recognition to generate or analyze the updated image.

107. Salah identifies eighteen “preferred methods” and eighteen methods that are “not preferred” but “can also be used,” for determining “discriminant information.” *Id.*, 34:15-36:14. Discriminant information includes, but is not limited to, contours. The preferred methods include long-established contour detection techniques, such as the Canny algorithm, which Salah specifically identifies as associated with determining contours. *Id.*, 34:15-16 (disclosing “well-known and preferred methods” including “applying a Canny filter, in particular to search for contours using Canny’s algorithm”). The non-preferred methods also include techniques that can be used in the process of determining and refining contours. *Id.*, 36:14. “Deep learning” is expressly identified at the bottom of the list

of methods that are “not preferred.” *Id.* As Dr. Foroosh agreed in his deposition, Salah is not teaching “deep learning” as a preferred method for determining any discriminant information, including contours, and has expressly identified more than a dozen other techniques that are “preferred” by comparison. EX2048,[Foroosh Tr.], 199:11-200:4.

108. Nowhere does Salah indicate that the contour map for the “updated image” would reflect identification of contours associated with individual identified teeth. To the contrary, the contour example in the definition of “discriminant information” contemplates that contour values are stored at the pixel level (0 and 1 indicating whether a pixel is part of a contour or not) and Salah does not mention any other potential values. EX1008, 18:14-15. Figure 6b is a simple contour map (*id.*, Fig. 6b), and while individual teeth may be identifiable in the contour map, there is no indication in Salah that they are identified during the analysis or annotated or otherwise marked on or associated with the “updated image.”

109. Salah’s reference to the disfavored approach of “deep learning” would not have suggested to a POSA using such a technique to identify which contours are associated with individual teeth, *i.e.*, semantic contours. This is not necessary for the purposes of Salah and Salah discloses that the analysis in “step c)” is “best carried out ... on the patient’s own device, preferably using the [image acquisition device]” (*i.e.*, a cellphone). *Id.*, 33:6-8. In light of this preference, a POSA—like the Salah

inventors—would have been discouraged a POSA from using a resource intensive “deep learning” approach to determine a contour map. *See, e.g., id.*, 33:6-8; 13:24-28 (“Preferably, steps (c), (d) and (e) are carried out ... ***without the use of any device other than a cell phone*** ... ”); 16:14-21 (“The invention also relates to: - a computer program, and in particular a specialized cell phone application, comprising program code instructions for execution of one or more, preferably all steps b) to f)... . in particular a cell phone or tablet, in which such a program is loaded.”).

110. In his deposition, Dr. Foroosh suggested that Salah Figure 11 reflects the determination of semantic contours or semantic segmentation because the annotation of dashes 20 and circles 22 on the updated image coincide with the patient’s teeth. *E.g.*, EX2054, 98:10-99:6 (“If you look at it, it’s only restricted to teeth ... that’s because ... obviously they have segmented out certain portions of the jaw[.]”); *id.*, 105:4-21 (“That discriminant information of reflection is restricted to teeth and appliances in teeth, and that indicates to a POSA that a segmentation has taken place. In other words, when Salah is disclosing a teaching that used contour detection algorithm to perform that segmentation, it clearly says that the ... extraction of discriminant information ...[is] using [that] contour detection and edge detection.”). In other words, Dr. Foroosh appears to suggest that an image segmentation algorithm (*e.g.*, deep learning) was used to recognize teeth in the

updated image and identify a contour bounding them. Dr. Foroosh misunderstands Salah, which discloses no such thing.

111. In Figure 11, a “reference map” of lines 20 and circles 22 from the computerized 3D reference model is “superimposed on the acquired image.” EX1008, 64:6-8. Those lines 20 and circles 22 represent “vectors” pointing in a direction “perpendicular to the faces [of the teeth] in the reference model.” *Id.*, 64:29-32. The lines 20 and circles 22 are not determined from any information in the acquired image—they come directly from the 3D reference model. *Id.* While this “reference map” almost perfectly overlaps with the patient’s teeth in the acquired image, that is not because the patient’s teeth were identified and segmented/contoured in the acquired image. Rather, the “reference map” directly overlaps with the patient’s teeth in Figure 11 because the location of the patient’s teeth are known *in the reference model*. See, e.g., *id.*, Fig. 4d. In other words, a POSA would have understood that Salah teaches pulling from the 3D reference model only those vectors associated with the faces of the patient’s teeth and “project[ing them] onto the reference image.” *Id.*, 64:29-32. As explained below, the “reference image” (from the 3D model) and the “acquired image” (from the patient) are aligned to reflect the same “acquisition conditions,” including the same perspective. Thus, the location of the patient’s teeth in the *reference* image coincides with the location of the patient’s teeth in the *acquired* image. If the

reference map of the patient's teeth from the *reference* image (lines 20 and circles 22) is superimposed on the *acquired* image, the map will necessarily coincide with the location of the patient's teeth, even though the acquired image was never analyzed to determine their contours. In fact, Salah does not disclose that any contours were determined in this embodiment *at all*—the embodiment described in connection with Figure 11 only discusses the use of “discriminant information” “relating to reflection information” rather than contours, as discussed elsewhere. *Id.*, 64:6-66:8.

112. Furthermore, the superimposition depicted in Figure 11 is not part of the process disclosed in Salah for aligning the acquired image with the reference image. A POSA would have understood it is a demonstrative figure that illustrates this embodiment of Salah, in which the brightest spots in the image (“those areas of the acquired image which have a brightness above a quality threshold”) can be expected to coincide with those areas of the teeth with surfaces directly facing the camera (represented by circles 22 in the reference image). *Id.* 64:13-65:14. As Salah describes, the process of aligning the reference image to the acquired image involves comparing the location of those circles 22 (the “reference *map*”) with the location of the brightest spots in the image (the “acquired *map*”)—*not* with the acquired *image* itself. *Id.* Thus, not only would a POSA have understood that Salah does not teach the semantic segmentation of teeth in the acquired image, a POSA

would also have understood that Salah does not teach creating the superimposed image depicted in Figure 11 as part of the image analysis.

113. Dr. Foroosh is thus incorrect that Salah discloses the determination of semantic contours, either in connection with Figure 11 or in any other instance.

4. Steps d) to f) – updating the 3D model

114. In the subsequent steps, Salah describes how this discriminant information from the “updated image” is ultimately used to update the 3D model to reflect any changes to the patient’s dental arch since the 3D model was originally formed or last updated. *Id.*, 36:17-47:24. In order to compare the “updated image”—a 2D photo—with the computerized 3D model, a corresponding “virtual” 2D image called a “reference image” is generated using computer algorithms to simulate what a photograph of the 3D “reference model” *would* look like if it could be photographed in the real world. *E.g., id.*, 40:7-12 (“The image acquisition apparatus is then virtually configured under the virtual acquisition conditions to be tested, in order to acquire a reference image of the reference model under these virtual acquisition conditions to be tested. The reference image thus corresponds to the image that the image acquisition apparatus would have taken if it had been positioned, relative to the reference model to be tested, and optionally calibrated in the virtual acquisition conditions to be tested”).

115. If the “reference image” is generated to simulate conditions closely analogous to those used to capture the “updated image”—conditions referred to as “*virtual* acquisition conditions”—it can be compared directly to the initial reference model, and their differences should reflect how the patient’s teeth have changed (*i.e.*, moved or deformed) since the time of the initial reference model. *E.g., id.*, 46:10-14 (“The final reference model corresponds substantially to the updated image. By comparing step f), it is possible to observe the differences between tooth positioning in step a) (initial reference model) and during acquisition of the updated image (step b)). In this way, it is possible to accurately determine the movements of individual teeth between these two stages.”). Notably, Salah discloses tracking orthodontic treatment only by analyzing tooth movements; it does not suggest assessing aligner fit.

116. To obtain a virtual “reference image” of the virtual 3D “reference model” that can be directly compared to the “updated image,” Salah first teaches determining “virtual acquisition conditions” that most closely reflect the actual acquisition conditions used to obtain the “updated image.” Salah defines “acquisition conditions” as “specify[ing] the position and orientation in space of an image acquisition apparatus relative to the patient’s teeth or a model of the patient’s teeth, and preferably the calibration of this image acquisition apparatus.” *Id.*, 17:11-13. “Calibration” is defined as consisting of “all the calibration parameter values,”

which are parameters “intrinsic to the acquisition apparatus” and “which affect[] the acquired image” such as “the depth of field,” “[e]xposure time,” “[f]ocal length,” and “[s]ensitivity.” *Id.*, 17:14-22. As an example, Salah depicts virtual images acquired at two different focal lengths in Figures 12a and 12b. *Id.*, 42:16-17, Fig. 12.

117. Optional “step d)” takes advantage of the presence of the “registration marks” in the “updated image” to approximate the acquisition conditions by evaluating their apparent alignment, orientation, size, and/or spacing. *Id.*, 36:17-38:9. The known orientation and geometry of the “registration marks” can be used to determine the positioning of the acquisition device relative to the patient’s teeth. *Id.* An example of the use of “registration marks” to determine orientation and distance is depicted in Figure 7. The size, distance, and orientation of the “registration marks” can be used to determine the angle and distance from which the photograph of the exemplary cube was acquired. *Id.*, 36:26-37:2.; Fig. 7. Such marks, when applied to the retractor (labeled as “12” in Figure 5a), can facilitate a similar determination with respect to the patient’s teeth. *Id.* The “registration marks” can also be used to evaluate focal length, exposure time, sensitivity, and depth of field. *Id.*, 38:19-39:2. The use of the “registration marks” for this determination “advantageously limits the number of tests on virtual acquisition conditions” in the subsequent processing and therefore “speeds up step e)

considerably.” *Id.*, 36:19-20. Some acquisition conditions such as “gyroscopic data,” can also be obtained directly from the acquisition device. *Id.*, 38:10-11.

118. In “step e),” the virtual acquisition conditions are determined, including by using the estimated conditions from “step d)” as a starting point, and the reference model is updated to reflect changes in the patient’s teeth. Salah describes an iterative process of comparing discriminant information such as contours between the reference images and updated image, including as the reference model is deformed to match the position and shape of the patient’s teeth as depicted in the updated image. *Id.*, 39:21-46:8.

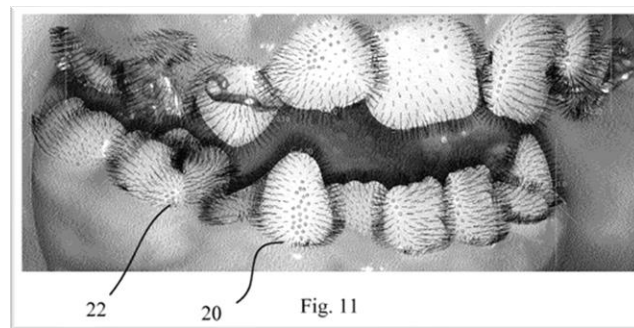
119. Finally, in “step f),” the final reference model can be compared with the initial reference model “to observe the differences between tooth positioning” and “accurately determine the movements of individual teeth between ... two stages.” *Id.*, 46:9-14.

120. It is my opinion that, as is apparent from the above description of Salah, the method for monitoring tooth displacement in Salah is distinct from the method for assessing the shape of an orthodontic aligner in the ’248 patent. Salah does not use images showing an aligner in a service position, and it nowhere suggests determining a value for a tooth or image attribute related to the separation between the aligner and the patient’s teeth, nor assessing aligner fit for treatment, as required by the challenged claims.

121. In addition, Salah's discussion of contours as "discriminant information" would not have suggested to a POSA that contours of individual teeth should be separately determined from the "updated image." Only the full contour of all teeth collectively needs to be obtained to determine the differences in outline between the reference image and the updated image. *See id.*

5. Salah's Alleged "Orthodontic Appliance"

122. Dr. Foroosh also relies on Salah as disclosing images of "teeth of a patient using orthodontic appliances." EX1002, ¶89. For this teaching, Dr. Foroosh points exclusively to Figure 11 of Salah (shown below), in which the patient is wearing a bonded expander on the upper teeth during imaging. *Id.*



123. As a preliminary matter, Salah does not disclose assessing the *distance* between the bonded expander shown in Figure 11 and the tooth, or any feature of the bonded expander. I also understand from Dr. Kusnoto that, unlike an aligner, this bonded expander on the upper teeth could not be readily removed by the patient. A POSA would have understood that the bonded expander on the upper teeth was included for that reason, not because it was desirable to acquire an image of the

bonded expander on the upper teeth. Salah would not have taught a POSA to acquire image of any removable device, such as an aligner.

124. Additionally, Salah's contour detection techniques are reliant on distinguishing between highly contrasting areas. For example, to assess tooth positions, Salah teaches that "[i]n the case of a contour, for example, contrast analysis provides more or less reliable information: an area of high contrast can be assimilated to an area corresponding to a contour with a high probability, and the quality of the points of this area will be high. On the other hand, a low-contrast area, such as a blurred zone, can be assimilated to an area corresponding to a contour with a low probability, and the quality of the points of this area will be low." EX1008, 60:17-22. Salah thus recommends processing the reference image to retain only the outlines identified with a high probability. EX2009, [0533] [U.S. Pat. Pub. Salah]; EX1008, 61:2-3. That is because, if the "quality threshold is low, the discriminant information retained will be abundant, but unreliable" and this "will then lead to the retention of 'false' contour points, i.e. points which, will be mistakenly considered as belonging to the contour of teeth and gums." EX1008, 60:25-61:1. Salah thus teaches that it is preferred that "the quality threshold is high in order to retain only highly reliable discriminating test information." *Id.*, 61:2-3. In other words, Salah teaches that high contrast is *required* to achieve a quality analysis. A POSA would have understood that an aligner would not provide a high contrast with the patient's

teeth or mouth because aligners visually blend with the teeth. Thus, a POSA would have understood Salah to teach away from using contour detection with aligner images as the methodology would need to be altered to detect areas of low contrast instead.

D. Maninis

125. Kevis-Kokitsi Maninis et al.’s article entitled “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks” (“Maninis,” EX1007), proposes a technique, which the authors refer to as Convolutional Oriented Boundaries (“COB”), for determining boundaries in an image using a convolutional neural network (“CNN”) (referred to in this declaration as the Maninis “COB neural network”), which can be applied in combination with other neural networks to determine semantic contours. EX1007, 1. In order to understand the disclosure of Maninis, I have included above a high-level explanation of neural networks as they would have been understood by a POSA in July 2017. *Supra* §VI.D. While that is a simple example, the Maninis COB neural network is much more complex and has more layers.

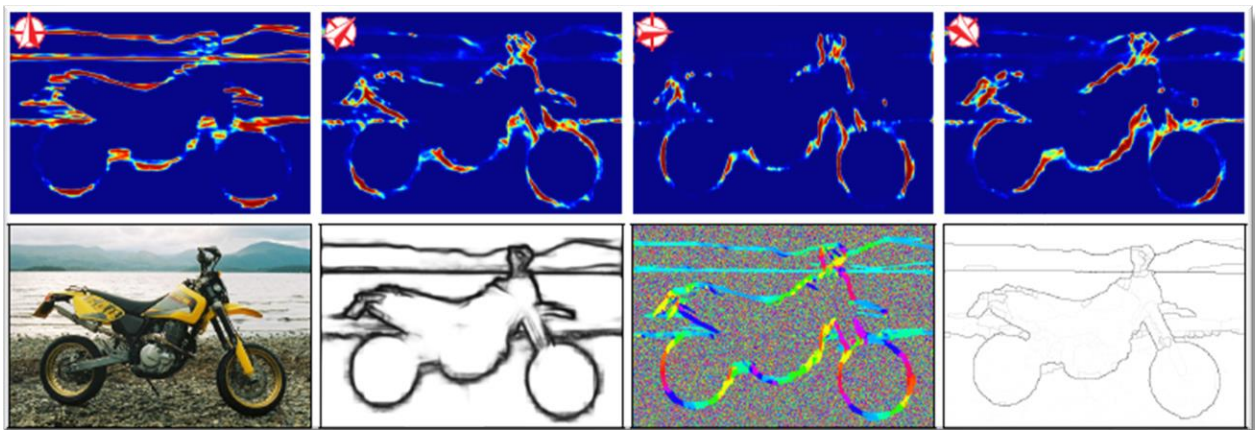
1. Overview of Maninis

126. Maninis is an academic paper and makes no mention of the use of its neural network architecture for analysis of dental or orthodontic images. Dr. Foroosh does not identify any reason why a POSA would have been aware

specifically of Maninis, particularly if searching in the field of dental imaging, and I am not aware of any such reason.

127. As noted above, Maninis proposes a technique, which the authors refer to as Convolutional Oriented Boundaries (“COB”), for determining boundaries in an image, using the Maninis COB neural network, which can be applied in combination with other neural networks to determine semantic contours. EX1007, 1.

128. In Maninis, the COB neural network identifies contours. *Id.*, 1-2. Maninis further discloses taking those contours and using them to form an ultrametric contour map (UCM) that is fused into a hierarchical segmentation. This hierarchical segmentation identifies boundaries that belong to the outline of an object (*e.g.*, a motorcycle) as opposed to a sub-part of that object (*e.g.*, the black region in the middle of the motorcycle’s seat), as shown in the example below:



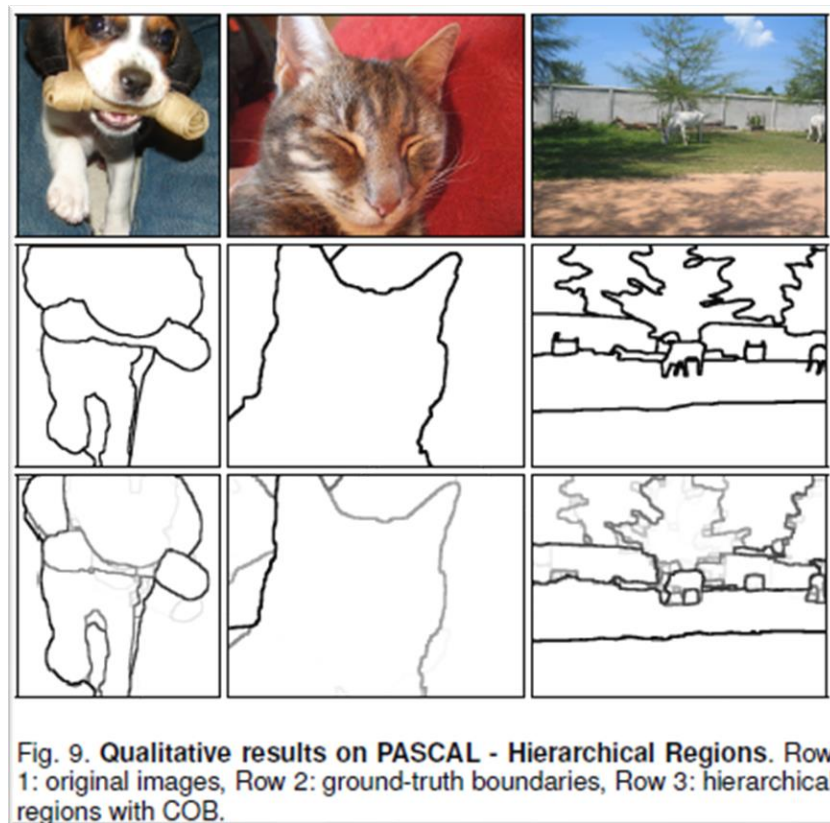
Id., 4, Fig. 3 (hierarchical regions depicted in bottom right image).

129. While Maninis’s COB neural network can identify the contours of objects depicted in images, its COB neural network **cannot** determine what object is defined by that boundary. This would have been clear to a POSA in July 2017 from the following disclosures of Maninis.

130. First, Maninis distinguishes between “Low Level Applications,” which are addressed in Section 5, and “High Level Applications,” which are addressed in Section 6. EX1007, 5-13. The low-level applications are those that can be achieved with Maninis’s COB neural network on its own, including “generic image segmentation” and “object boundary detection.” *Id.*, 5-9. A POSA would have known that “generic image segmentation,” refers to the partitioning of an image into parts corresponding to objects, but as the name suggests, it does not identify what those objects are. As Maninis explains, with “object boundary detection,” “[t]he goal is to detect the boundaries of the objects that belong to the 20 classes of PASCAL, **without distinguishing the semantics.**” EX1007, 7. A POSA would have understood that “without distinguishing the semantics” means without identifying what object is contained within the boundary.⁴ This is illustrated by Figures 9 and

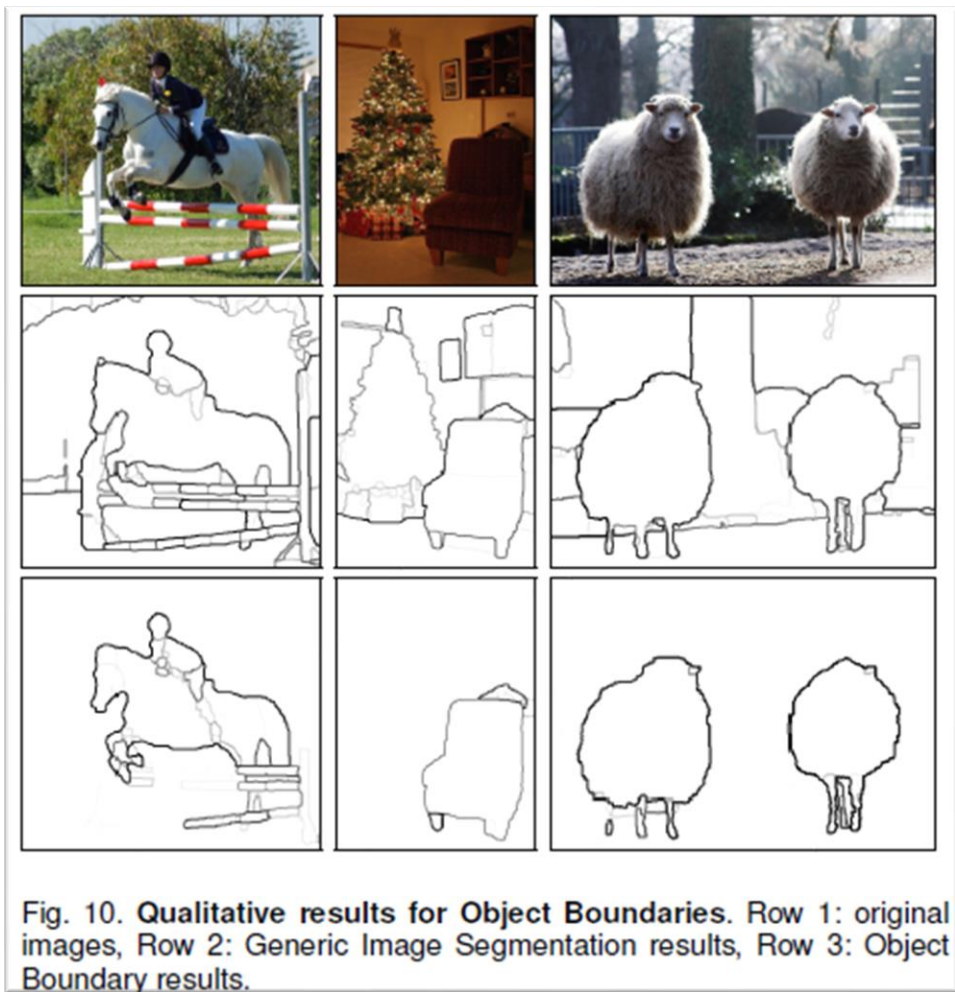
⁴ As of 2017, a POSA would have understood that “boundaries” or “contours” are the outlines of objects depicted in an image. A POSA would have known that contour (or edge) detection is the process of detecting the location of those contours

10. Figure 9 shows the results of Maninis’s “generic image segmentation” with row 3 depicting the hierarchical boundaries Maninis determined using the results of its COB neural network. EX1007, 7, 9, Fig. 9 (below).



(or edges) in an image. In practical terms, this means identifying whether each pixel in an image is part of a contour or not. A segment is the region within a contour. Similarly, a POSA would have known that “semantic segmentation” is the labeling of pixels in an image as associated with a particular type of object and “semantic boundaries” are the contours (or edges) that define a particular type of identified object.

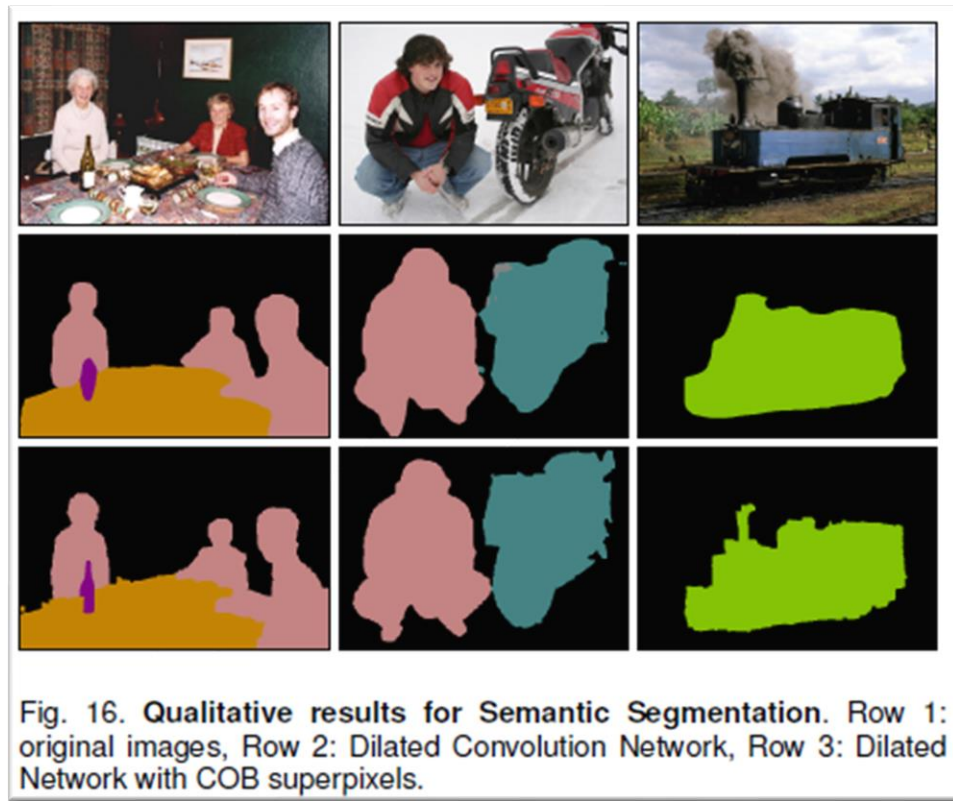
Although Maninis describes these contours as “semantically meaningful,” a POSA would have recognized that no “semantic segmentation” or object classification was performed by the Maninis COB neural network. EX1007, 7. Like Figure 9, Figure 10 illustrates “generic segmentation and object boundary detection,” but not semantic segmentation or object identification. EX1007, 7-9, Fig. 10 (below). As shown, Maninis’s COB neural network detects only contours, specifically generic image segments (second row) and object boundaries (bottom row):



EX1007, 9, Fig. 10. Although Maninis’s COB neural network identifies the contours of objects in these images, the COB neural network is not able to determine the specific type of object that these contours bound (*e.g.*, a horse, a chair, or a sheep). *See id.* This contrasts with Figure 16 (reproduced below), produced by a separate neural network capable of semantic segmentation, where each region has a color identifying it as background or as associated with a particular object (*e.g.*, the color green for a train) (discussed further below).

131. Second, in order to achieve “High Level Applications,” including “semantic segmentation” and “object detection,” the Maninis authors expressly indicate that it is necessary to use other techniques—specifically, they use other neural networks described by other authors in other publications. EX1007, 9-13; *id.*, 13 (stating that the Maninis COB neural network was used “***in combination with***” other neural networks that perform the “high-level vision tasks (object proposals, object detection, and semantic contours and segmentation).”). Specifically, Maninis discloses that to determine the “***semantic*** contours” on which Dr. Foroosh relies for purposes of his proposed obviousness combination, the COB neural network was combined with one of two additional neural networks (“Dilated Network” and “PSPNet”). *See id.*, 9 (“In Section 6.2 we study the interplay of our boundary detector with semantic contours and semantic *segmentation by combining COB with Dilated Network [68] and PSPNet [69][.]*”). Maninis further expressly discloses

that they “couple[d] the COB boundaries with Semantic Segmentation results” from the other neural networks by overlapping them (as a “mask”) or by forcing the semantic regions to align with the nearest COB boundaries (“snapping”), as pictured in Figure 16 (below):



Id., 11-12, Fig. 16. The second row shows the results of the Dilated Convolutional Network, a neural network described in a separate publication that Maninis cites. The third row represents the result of “[s]napping” the semantic segmentations determined by the Dilated Convolutional Network to the contours determined by the Maninis COB neural network. In other words, the Maninis COB neural network

only refined the definition of the segment boundaries, it did not itself identify what was depicted within each segment.

132. The combination of the Maninis COB neural network with other neural networks for determining semantic contours is also reflected in Tables 4-6, where object detection results of “COB-dil” and “COB-PSP” (the combinations of the Maninis COB neural network with the “Dilated Convolution network” and the “PSPNet” network, respectively) are compared against other prior art techniques for determining object “class” (*e.g.*, “Plane,” “Bicycle,” “Bird,” etc.). *Id.*, 11-12, Tables 4-6.

133. Thus, to determine “semantic boundaries,” which “requires not only detecting the boundaries, but also associating a semantic class to them,” a POSA would have recognized the need to employ **both** the Maninis COB neural network, which only provides “binary information of boundaries,” **and** a separate neural network capable of semantic segmentation, which “label[s] each of the detected pixels with the corresponding semantic class,” and then “fuse the results of the two tasks.” *Id.*, 11.

134. A POSA would have also understood that Maninis discloses using a second neural network (“Fast-RCNN”) in connection with the Maninis COB neural network to perform “object detection,” as described in Section 6.3. *Id.*, 9-13. In particular, a POSA would have understood that Maninis discloses taking the COB

neural network contours, using them to generate a “bounding box”—*i.e.*, a box surrounding the part of the image in which the contours indicate the presence of an unidentified object—then submitting the image with that “bounding box proposal” to the Fast-RCNN neural network. *Id.*, 12 (“We evaluate the bounding box proposals generated by COB by feeding them into the Fast-RCNN [70] pipeline for Object Detection.”). It is the Fast-RCNN neural network that determines a probability as to the class of object that appears in the box (*e.g.*, “cat”). *Id.* Table 7 compares the results for object detection by class (*e.g.*, for detecting a “Plane,” “Bicycle,” “Bird,” etc.) *using Fast-RCNN* when starting with “bounding box proposals” from either the COB neural network results or the “Selective Search [60] algorithm” (“SeSe”). *Id.*, 12, Table 7 (below) (showing “*performance ... of Fast-RCNN ... using object proposals from [Selective Search] (original) or COB*”).

Technique	Plane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	MBike	Person	Plant	Sheep	Sofa	Train	TV	Mean
COB	69.5	76.8	69.7	53.3	44.6	80.5	81.3	83.1	45.3	74.2	69.4	80.1	84.2	76.7	72.8	35.9	67.1	68.4	75.1	65.4	68.7
SeSe	76.0	76.8	65.3	54.6	38.0	76.5	78.2	81.6	40.1	74.1	66.5	78.9	81.8	74.5	66.2	32.9	65.6	67.7	73.4	66.8	66.8

TABLE 7
VOC 2007 test evaluation: Object Detection performance (mAP) of Fast-RCNN [70], using object proposals from [60] (original) or COB.

135. Third, the fact that the Maninis COB neural network determines only generic boundaries is also confirmed by how the Maninis COB neural network is trained. In particular, the Maninis COB neural network is trained with other publicly available image sets that were formed for image recognition algorithm competitions, such as the “PASCAL Context dataset.” *Id.*, 5-6. The PASCAL Context dataset

consists of more than 10,000 images that have been labeled with “carefully localized pixel-wise semantic annotations” corresponding to “459 semantic categories.” *Id.*, 5-6. While the PASCAL Context dataset contains regions that are labeled as corresponding to identified objects (*i.e.*, “semantic segmentations”), the Maninis COB neural network is trained **only** on the contours themselves (easily derived from the segmentation) without the semantic information identifying the objects those contours represent. This is demonstrated when the Maninis authors discuss the training set in mathematical terms in Equation 1 below:

linearly to produce a contour probability map. Using the notation of the authors, we denote the training dataset by $S = \{(X_n, Y_n), n = 1, \dots, N\}$, with X_n being the input image and $Y_n = \{y_j^{(n)}, j = 1, \dots, |X_n|\}$, $y_j^{(n)} \in \{0, 1\}$ the predicted pixelwise labels. For simplicity, we drop the subscript n . Each of the M side outputs minimizes the objective function:

$$\ell_{side}^{(m)}(\mathbf{W}, \mathbf{w}^{(m)}) = -\beta \sum_{j \in Y_+} \log P(y_j = 1 | X; \mathbf{W}, \mathbf{w}^{(m)}) - (1 - \beta) \sum_{j \in Y_-} \log P(y_j = 0 | X; \mathbf{W}, \mathbf{w}^{(m)}) \quad (1)$$

where $\ell_{side}^{(m)}$ is the loss function for scale $m \in \{1, \dots, M\}$, $\mathbf{W}$ denotes the standard set of parameters of the CNN, and $\{\mathbf{w}^{(m)}, m = 1, \dots, M\}$ the corresponding weights of the the m -th side output. The multiplier β is used to

Id., 3. Specifically, the “predicted pixelwise labels” (*i.e.*, “ Y_n ”) are identified as a dataset consisting of values selected from the set of zero and 1 (*i.e.*, “ $\{0, 1\}$ ”). *Id.* Because the only labels for the pixels in the training images are “0” (not part of a contour) and “1” (part of a contour), the Maninis COB neural network is never taught

to associate those contours with any particular object (*e.g.*, a cat), which would require additional value options. This is consistent with the paper Maninis cites for the above formula, which identifies this formula as corresponding to a “ground truth **binary** edge map.” See EX2042,[Xie], 3 (identified as reference [2] in EX1007). Maninis indicates that Equation 1 describes how the Maninis COB neural network functions for the entirety of the paper. EX1007, 4 (“In the rest of the paper we use the class-balancing cross-entropy loss function of Equation 1.”). This is further confirmed by Section 5.4 where Maninis states that they “extract[ed] contours from the semantic segmentation annotations of the [VOC] database” to train their COB neural network for object boundary detection. *Id.*, 7.

136. Accordingly, when Maninis reports the “Efficiency Analysis” in Section 5.6, including the “Total Time” (*i.e.*, 0.79 seconds) it took to determine hierarchical regions (“OWT, UCM”) in an image, a POSA would have understood that these times *only* account for generic image segmentation. *Id.*, 9, Table 3. A POSA would have further understood that Maninis does not disclose how long it took to determine semantic contours or recognize objects (*e.g.*, cat, sheep, motorcycle) using the other two neural networks that Maninis used to perform that analysis (or any other neural network). Dr. Foroosh does not identify any evidence related to the time it would take to run the Maninis COB neural network together with any of the other techniques Maninis identifies as necessary to achieve semantic

contours or object recognition, under any conditions (*e.g.*, with any particular hardware).

137. Furthermore, Maninis reports its COB neural network can be used to obtain generic image segmentation in 0.79 seconds when computed “on a Nvidia Titan X GPU.” An Nvidia TITAN X GPU has 12 GB of memory and a processing capacity of 11 TFLOPS. EX2045,[Nvidia TITAN X]. Floating point operations per second or “FLOPS” is an indicator of the rate at which a processor can make individual calculations. A capacity of 11 TFLOPS indicates that the Nvidia Titan X was capable of performing 11 trillion floating point operations per second. A POSA would have understood that the capability of a Graphics Processing Unit (GPU) to handle many calculations simultaneously is critical for running the Maninis COB neural network, which requires billions if not trillions of individual calculations to analyze images. For this reason, as Dr. Foroosh identified in his declaration, “[d]evelopments of parallel processing hardware (in particular Graphic Processing Units) made it possible for deep learning networks to be trained in an end-to-end manner and on much larger training sets ... involving tens of thousands to millions of images.” EX1002, ¶31.

2. Training a Neural Network Requires a “Learning Base”

138. Neural networks as disclosed in Maninis, including both Maninis’s COB neural network and the other neural networks for semantic contours and object

detection from other authors, must be trained on a data set or “learning base”—a set of images labeled with the information the neural network is being trained to identify. In Dr. Foroosh’s proposed combination, the “learning base” comprises at least 10,000 dental images of aligners in a service position, acquired with cellphones, with pixel-wise labels of teeth, aligner, and the separation between them. EX1002, ¶122; EX2048, 103:23-106:5, 222:7-12 (confirming testimony regarding motivation to use 10,000 image learning base discussed in connection with claim 5 applies to his combination for all claims).

139. A significant driver for the rising popularity of the application of deep learning devices to problems in other fields (outside of dentistry and orthodontics⁵) was the availability of large, pre-existing databases of annotated images. *E.g.*, EX1002, ¶31; EX2041,[Russakovsky] (cited as reference [7] in Maninis) (identifying database comprising over a million labeled images publicly available for experimentation). However, Dr. Foroosh fails to identify any collection of at least 10,000 aligner images, acquired with cellphones, labeled with pixel-wise labels of teeth, aligner, and the separation between them, and I am aware of none. Nor

⁵ Dr. Foroosh did not identify any prior art example of the application of deep learning in the dental field, and specifically, to the analysis of dental photographs showing an aligner in a service position, and I am not aware of one.

does Dr. Foroosh cite a single example of a neural network trained on such aligner images, and I am also aware of none. None of the images or categories in the datasets used to train the Maninis COB neural network include aligner images or teeth. EX2048, 212:16-214:5. In fact, an article from 2021 notes the lack of pre-existing databases of dental images for training neural networks. EX2010,[Liu], 1072 (Even in 2021, “[c]ompared to other medical fields, the high-quality datasets for orthodontic research [were] limited. The discrepancies in training data make the comparison of different AI-based models questionable. Although supervised learning is currently the optimal choice in malocclusion diagnosis, huge costs and labor required for target labeling create obstacles in creating standardized and high-quality datasets for orthodontic studies, specifically.”).

140. I understand from Dr. Kusnoto that it was not a regular practice in 2017 for dental practitioners to acquire images of seated aligners with mobile devices and he is not aware of any databases of such images available as of July 2017. Dr. Kusnoto further informed me that, there would be a significant expense associated with labeling photographs to create the learning base and also with de-identifying patient data for purposes of privacy laws such as HIPAA, which would be required before they could be used in a learning base. Accordingly, a POSA seeking a learning base of aligner images with pixel-wise labels of teeth, aligner, and the

separation between them, as proposed by Dr. Foroosh, would have recognized the significant time and cost associated with creating such a learning base.

VII. CLAIM CONSTRUCTION

A. Constructions Proposed by Dr. Foroosh

141. Dr. Foroosh proposes constructions of the claim terms “tooth attribute,” “image attribute,” and “image extracted from a film.” Ex. 1002, ¶¶64-68. I address each in turn below.

1. “tooth attribute”

142. Challenged claims 1, 2, 7, and 14 recite the term “tooth attribute.” Dr. Foroosh opines that “tooth attribute” “refers to any attribute related to a tooth” and “includes any attribute of a tooth that may relate to the tooth’s separation from an aligner, including attributes specific to the tooth or to the aligner, such as their respective positions or edges, which may bear a relation to the separation between the tooth and the aligner.” *Id.*, ¶¶64-65.

143. I disagree that “tooth attribute” can be construed so broadly as to include the “respective positions or edges” of the tooth and the aligner. In my opinion, Dr. Foroosh is incorrect that the position or edge of the tooth, on its own, can be a tooth attribute as that term is used in the ’248 patent.

144. The ’248 patent describes a detailed analysis method that uses a deep learning device trained on a learning base comprising historical images of dental

arches representing an aligner on a historical patient “comprising one or more zones each representing a tooth, or ‘historical tooth zones,’ to each of which is assigned, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, **a tooth attribute value.**” EX1001, 28:62-29:13. This method for monitoring an orthodontic aligner further comprises “submission of the analysis image to the deep learning device for the deep learning device to determine at least one probability relating to the presence, at a location in said analysis image, of an analysis tooth zone, and the **tooth attribute value** of the tooth represented on said analysis tooth zone,” and “determination, as a function of said probability, of the presence of a separation between the aligner and the tooth represented by said analysis tooth zone, and/or of an amplitude of said separation.” *Id.*, 29:16-26; *see also id.*, 15:57-16:6, 16:37-54, 30:35-53. The ’248 patent thus distinguishes between a “tooth zone,” which corresponds to the “zone” or “location” of a tooth in the analysis image, and a “tooth attribute” which is an “attribute” of the tooth represented “on said ... tooth zone.”

145. The ’248 patent teaches that:

The tooth attribute is preferably chosen from a tooth number, a tooth type, a tooth shape parameter, for example a tooth width, in particular a mesio-palatine width, a thickness, a crown height, an index of mesial and distal deflection of the incisive margin, or a level of abrasion, a tooth appearance parameter, in particular a translucency index or a color parameter, a parameter relating to the condition of the tooth, for example “abraded”, “broken”, “decayed” or “fitted” (that is to say in

contact with a dental, preferably orthodontic, appliance), an age for the patient, or a combination of these attributes.

Id., 7:64-8:7. Each of these is a quality of the tooth, derived for an analysis tooth zone that represents (at least partially) a tooth, using a deep learning device. *Id.*, *e.g.*, 16:65-17:4 (It is “possible to establish a probability threshold such that, if a probability associated with an analysis tooth zone and with a tooth attribute value for this analysis tooth zone exceeds said threshold, ***the analysis tooth zone effectively represents a tooth having said tooth attribute value.***”). Moreover, each of the tooth attributes set forth in the specification of the ’248 patent is capable of being expressed as a single value (*i.e.*, a single numerical value such as tooth number or tooth width, or a single descriptive parameter such as tooth type, “abraded,” or “broken”). *Id.* 7:64-8:7; *see also id.* 8:12-16 (“For example, the ‘tooth type’ tooth attribute will have the value “incisor”, ‘canine” or ‘molar’ depending on whether the tooth model is that of an incisor, of a canine or of a molar, respectively.”).

146. In his discussion of the specific claim limitations, Dr. Foroosh opines that “the locations of the surfaces of the tooth and aligner, which would correspond to the detected contours” are tooth attributes. EX1002, ¶107. That is, Dr. Foroosh says that contours of regions detected according to Maninis that contain teeth are both (1) the outline of the claimed analysis tooth zone, and (2) the claimed tooth attribute. That reading is not supported by the ’248 patent.

147. As an initial matter, unlike the exemplary tooth attributes of the '248 patent, the contour of a region cannot be expressed as a single value. It would instead necessarily be expressed by a vector of values for each pixel in the image, each value representing whether that pixel is part of a tooth or aligner boundary line. In the exemplary neural network discussed above, there would be 784 such values. *Supra* ¶66. A POSA would not have understood the tooth attribute of the '248 patent to include qualities that can only be expressed as a complex *set* of values, none of which individually describe any discernable attribute of a tooth.

148. Moreover, if Dr. Foroosh is correct that the contour defining the analysis region is both the tooth attribute and the “analysis tooth zone,” these two features of the claims would merge into one. That is, the contour would both define the analysis tooth zone and be the tooth attribute. However, the '248 patent teaches, and the challenged claims recite, identifying and determining a value for a tooth attribute *of* an analysis tooth zone. For example, the '248 patent discloses determining the location of an analysis tooth zone and “the tooth attribute value of the tooth represented on said analysis tooth zone”:

b"3) submission of the analysis image to the deep learning device for the deep learning device to determine at least one probability relating to
the presence, at a location of said analysis image, of an analysis tooth zone, and
the tooth attribute value of the tooth represented on said analysis tooth zone.

EX1001, 29:16-22. Challenged claim 1 likewise recites “b”) analysis of the analysis image by means of a deep learning device, trained by means of a learning base, ***so as to determine a value for at least one tooth attribute of an ‘analysis tooth zone’ representing, at least partially, a tooth on said analysis image***, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image.” Accordingly, it cannot be that the tooth attribute is the analysis tooth zone itself, nor does the specification ever suggest as much. I address this further in the context of claims 7 and 14 below. *Infra* §VIII.D.3.

149. As such, it is my view that the proper meaning of “tooth attribute” is “attribute of a tooth,” and that it does not include the respective positions or edges of the tooth and aligner, as Dr. Foroosh contends.

2. “image attribute”

150. Challenged claims 8 and 15-16 recite the term “image attribute.” I agree with Dr. Foroosh that “[t]he ’248 patent describes an image attribute as an attribute that has a value that ‘relates to the whole image and not to part of the image.’” EX1002, ¶66 (citing EX1001, 3:23-37). I also agree “that an image attribute may be ‘determined as a function of the tooth attribute values.’” *Id.*, ¶67 (citing EX1001, 26:3-14). However, as set forth above, Dr. Foroosh’s construction of “tooth attribute” is overly broad, and thus so too is his interpretation of “image

attribute.” Dr. Foroosh also asserts that, in the context of claim 8, the “locations of each tooth and each corresponding aligner part would also be image attributes.” *Id.* ¶140. But as described above regarding “tooth attributes,” the contours of objects in an image are not “attributes,” they are “zones.” Nor do those locations or contours “relate[] to the whole image,” and are thus not “**image** attributes.” While an attribute of the whole image may be “determined **as a function of**” the contours of objects (*e.g.*, by deriving additional features in subsequent analysis, such as determining an average separation between the aligner and all teeth in the image), that does not make contours “image attributes” themselves. I address this further in the context of claim 8 below. *Infra* §VIII.D.4.

151. As such, it is my view that the proper meaning of “image attribute” is any attribute related to the image as a whole, and that it does not include the respective positions or edges of the tooth and aligner, as Dr. Foroosh contends.

3. “image extracted from a film”

152. Challenged claims 1 and 15 recite the term “image extracted from a film.” Dr. Foroosh opines that this term means “an image extracted from a video recording.” EX1002, ¶68. The construction of this term does not change my opinions set forth herein, and accordingly, I have applied Dr. Foroosh’s proposed construction.

B. Additional Constructions

1. “deep learning device”

153. The term “deep learning device” appears in challenged claims 1, 7, 8, and 13-16. Dr. Foroosh does not set forth any construction for deep learning device, however, he states he was instructed that the phrase “analysis of the analysis image by means of a deep learning device trained by means of a learning base” “should be interpreted to encompass a device comprising a deep learning neural network trained with a learning base, performing a function of analyzing an image (*i.e.*, the analysis image).” EX1002, ¶¶93-94.

154. In my opinion, a POSA would have understood that, in the context of the '248 patent, the term “deep learning device” refers to a device that, through training, is capable of analyzing an image and recognizing patterns therein. EX1001, 16:46-48 (“Through its training in the step 2), the deep learning device is capable of analyzing the analysis image and of recognizing said patterns therein.”); *see also id.*, 17:30-32 (same), 24:26-47. I understand that Dr. Foroosh generally agreed with this interpretation at his deposition. EX2054, 169:22-171:11 (“[G]enerally speaking, deep learning needs to be either trained or be a pretrained model to perform the task that you’re looking for. That task might be anything. It could be recognizing a pattern, but it could be many other things.”). While I agree that a neural network is

an example of a deep learning device, I disagree that “*device*” refers to a physical device (*e.g.*, processor or computer) in the ’248 patent.

155. The ’248 patent discloses that the invention includes “a computer program, and in particular a dedicated application for cellphones, comprising program code instructions for the execution of one or more steps of any method according to the invention, when said program is run by a computer.” EX1001, 31:55-60. It further repeatedly states that the “deep learning device” is “preferably a neural network.” *E.g., id.*, 1:22-28, 1:39-40, 1:66-67, 3:14-18, 8:31-33; 13:54-55, 16:5-35. The ’248 patent expressly identifies that “[a] ‘neural network’ or ‘artificial neural network’ is a set of **algorithms** well known to a person [of] skill[] in the art,” identifying several well-known neural networks as examples. *Id.*, 16:5-45. The ’248 patent thus confirms that the “deep learning device” is preferably an “algorithm,” which is a program written in computer code. Accordingly, a POSA would have understood that the word “device” in “deep learning device” refers to a computer program or algorithm(s), not to the hardware on which the algorithm runs.

156. This is consistent with a POSA’s understanding that a neural network is a model or algorithm represented in computer code. While a neural network runs on computers and other electronic devices, it has no physical structure of its own. *See, e.g., id.*, 31:55-60 (the invention includes “a computer program, and in particular **a dedicated application for cellphones**, comprising program code

instructions for the execution of one or more steps of any method according to the invention, when said program is run by a computer.”). Dr. Foroosh agrees that a “neural network” is an algorithm or model written in code, not a physical device. In his deposition, Dr. Foroosh provided a similar description of a neural network, clarifying while describing neural network “architecture”—specifically the deep “layers” through which the analysis and training data flow—that he was “not talking about hardware” and confirming that a neural network can be described as “an algorithm with values that are updated through the training process.” EX2048, 201-204; 209:24-210:5 (confirming that, “once ... trained,” neural networks can be “put ... on any number of computers and run[.]”); *see also* EX2054, 121:8-11 (“[Y]ou keep saying this learning device. That’s a somewhat unusual term. I would just refer to it as ‘deep learning’”).

2. “shape of an orthodontic aligner”

157. The term “shape of an orthodontic aligner” appears in the preamble to claims 1 and 15 of the ’248 patent. I am informed by counsel that whether a claim preamble is a substantive limitation is determined based on the overall form of the claim and the invention as described in the specification and in the prosecution history. I understand that a preamble is limiting when it is necessary to understand the subject matter encompassed by the claim, when it provides an antecedent basis to terms used in the body of the claim, when it presents the essence or a fundamental

characteristic of the invention, when patentability depends on the limitations stated in the preamble, or when the preamble recites structure underscored as important by the patentee. By contrast, a preamble is less likely to be limiting where it merely gives a name or describes a use or purpose of the invention.

158. I understand that it is Dental Monitoring's position that, in the event that the preamble is limiting, the "shape of an orthodontic aligner" should be construed to mean "the shape of the aligner relative to the shape of the teeth of the patient." In my opinion, this construction is consistent with the specification of the '248 patent, and I agree with this construction. For example, the '248 patent relates to a method of "assessing the shape of an orthodontic aligner" for purposes of aligner treatment. *See, e.g.*, EX1001, claim 1; *see also id.* 28:35-37 (describing the invention as "considerably simplify[ing] the assessment of the good suitability of the aligner to the treatment"), 30:60-62 ("The suitability of the aligner can be assessed in the context of an orthodontic treatment (separation [*i.e.*, space between the aligner and the teeth] compatible or not with the orthodontic treatment)").

159. The '248 patent further explains that the shape of the aligner is assessed relative to the patient's teeth so as to determine its suitability for treatment. EX1001, 4:54-62 ("It defines a generally 'U' shaped channel, the shape of which is determined to ensure the fixing of the aligner onto the teeth, but also as a function of a desired target positioning for the teeth. More specifically, the shape is

determined in such a way that, when the aligner is in its service position, it exerts stresses tending to move the teeth being treated to their target positioning, or to hold the teeth in this target positioning.”). A POSA would have understood that, as set forth in the ’248 patent, the “shape of an orthodontic aligner” can be assessed by determining “a value for at least one tooth attribute of an ‘analysis tooth zone’ of the analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image.” EX1001, 28:10-33; *see also id.*, 7:4-6, Fig. 9 (“FIG. 9 represents, schematically, the different steps of a method for assessing the shape of an orthodontic aligner of a patient, according to the invention,” which includes the steps recited in claim 1 of the ’248 patent and includes an “assessment of the suitability of the aligner to the patient”). Indeed, the ’248 patent expressly provides that the “value for this tooth attribute therefore provides an information item relating to the shape of the aligner relative to the shape of the teeth of the patient.” *Id.*, 30:42-45.

160. Dr. Foroosh opines that the preamble is satisfied by alleged prior disclosure of a method of “assessing the fit of the aligner” that involved “determining a distance between the teeth and the aligner when worn by the patient.” EX1002, ¶84. Dr. Foroosh thus appears to understand that the preamble, if limiting, should be construed to refer to a method of assessing “the shape of the aligner relative to the shape of the teeth of the patient,” *i.e.*, aligner fit.

VIII. THE '248 PATENT WOULD NOT HAVE BEEN OBVIOUS IN VIEW OF KIM, INVISALIGN, SALAH, AND MANINIS

161. For the reasons set forth below, I disagree with Dr. Foroosh that the challenged claims of the '248 patent would have been obvious in view of Kim, Invisalign, Salah, and Maninis.

162. In reaching these opinions, I have relied on my discussion in each of the preceding sections as further explained below.

A. Dr. Foroosh's Proposed Combination

163. Dr. Foroosh's combination relies on all four of Kim, Invisalign, Salah, and Maninis. I understand that Petitioner identifies Kim as the primary reference and Invisalign, Salah, and Maninis as secondary references. Paper 2, 14, 20. However, Dr. Foroosh does not appear to adopt that distinction. My opinion does not change regardless of which reference is identified as the primary reference.

164. Dr. Foroosh's analysis begins with Kim. EX1002, ¶69. According to Dr. Foroosh, "Kim discloses methods for providing remote tooth alignment services, including evaluation of a clear aligner to be worn by a patient and determination of, for example, 'when a new or modified clear aligner is necessary.'" *Id.* (quoting EX1005, Abstract, ¶10). Dr. Foroosh further states that "Kim also teaches that automatic recognition of contours of a patient's face and teeth may be employed as part of the imaging process," pointing to Kim's disclosure of displaying a virtual tooth line on-screen during photo capture. EX1002, ¶69 (citing EX1005, [0014]).

Dr. Foroosh does not suggest that this aspect of Kim would have motivated a POSA to assess the fit of an aligner by determining contours. I agree that Kim would not have provided such a motivation because this teaching concerns the acquisition of images of the patient's teeth, without an aligner in place. As discussed above, Kim does not suggest any non-human analysis of orthodontic treatment from the acquired images. Instead, Kim is concerned with the assessment of teeth alignment using images of a patient's teeth without aligners present; this is a different assessment from aligner fit or the shape of an aligner.

165. Dr. Foroosh then points to the clear aligner information transmission program of Kim, as “operat[ing] to provide orthodontists with images of aligners worn by the patient that may be usefully analyzed by the orthodontist,” and that this analysis “reduce[s] diagnosis time and cost” as well as “enhanc[ing] accuracy.” EX1002, ¶70 (citing EX1005, [0008]-[0010], [0066], [0072]). As explained above, however, the clear aligner information transmission program of Kim involves only photographs of the aligners *on their own*. See *supra*, §VI.A. It is the “teeth alignment information”—referring to information about the patient's teeth without an aligner in place—that is used for diagnosis and to determine whether new aligners need to be ordered, not the “clear aligner information,” which refers to information about the aligner on its own. *Id.* Accordingly, the alleged benefits of Kim come from the capture and review of images of the teeth, not the aligner as on the patient's

teeth. Thus, it is my opinion that Dr. Foroosh's motivation to combine the references is based on an incorrect understanding of Kim. Moreover, Dr. Foroosh does not articulate *any* other motivation for combining Kim with the other references beside his mistaken reading of Kim.

166. Dr. Foroosh next alleges that "Invisalign corroborates Kim's teachings regarding imaging an aligner in a service position on teeth." EX1002, ¶71. As explained above, Kim nowhere discloses photographing or otherwise imaging an aligner seated on a patient's teeth. Nonetheless, Dr. Foroosh goes onto allege that Invisalign teaches "assessments that can be used to determine whether the clear aligner is normal and corresponds to an appropriate step," including "that if an aligner is well seated, there will be a distance of less than 1 mm between the end of the teeth and the surface of the aligner." *Id.* Dr. Foroosh points to one image of a poorly-seated aligner in Invisalign to demonstrate that the resulting gap of more than 1 mm "is plainly visible on the photograph." *Id.* But Invisalign is concerned with an analysis that is performed in person, it instructs orthodontists to perform certain steps at each appointment that could not be performed remotely, and I understand from Dr. Kusnoto that the image on which Dr. Foroosh relies is not representative of the typical aligner fit observed by dental practitioners at the time. *See supra*, §VI.B.

167. Dr. Foroosh then asserts that “[i]n view of Kim’s disclosure of orthodontic checking of a photo or video of an aligner currently worn by the patient and Invisalign’s teachings of how to evaluate aligner fit, illustrated by an image of an aligner worn on a patient’s teeth, a POSA would have had reason to analyze images of the teeth of a patient wearing an aligner in a service position to evaluate whether the aligner fits appropriately.” EX1002, ¶72. As noted above, Dr. Foroosh’s opinion is based on a misunderstanding of Kim’s and Invisalign’s disclosures—neither Kim nor Invisalign suggests evaluating aligner fit from an image of a patient wearing an aligner. *See supra*, §§VI.A., VI.B. Dr. Foroosh is thus incorrect that the combination of Kim and Invisalign “would be applying known methods for orthodontic evaluation of aligner fit.”

168. Dr. Foroosh concludes that a POSA would thus have reason to evaluate images for aligner fit, and thus separation, including because Kim teaches that remote treatment results in enhanced accuracy and more efficient results. *Id.* ¶72 (citing EX1005, [0010], [0072]). But again, those alleged benefits of Kim stem from the capture and review of images of the teeth, not the aligner on the patient’s teeth. *See supra*, §VI.A.

169. Dr. Foroosh next points to the reference Frey⁶ (EX1009) as “confirm[ing]” “that POSAs were motivated” to use images of aligners for orthodontic evaluation. Frey describes a “DIY” method of implementing Dental Monitoring’s software and does not relate at all to using computer or deep learning analysis of images. EX1009, 17-27. Instead, Frey suggests obtaining a video which the orthodontist would then remotely visually inspect.

170. As none of Kim, Invisalign, Salah, or Frey disclose determining the width of gaps between teeth and aligners from a photograph, Dr. Foroosh pivots to argue that “[a] POSA would have known how to determine the width of such gaps using image processing software.” EX1002, ¶74. Dr. Foroosh argues that Salah provides useful disclosure by teaching the determination of the edges of objects in cellphone pictures of teeth using deep learning. *Id.* ¶74. He further opines that, “[l]ike Kim, Salah teaches application of computer image analysis to dental images for orthodontic treatment purposes.” *Id.* I disagree that Dr. Foroosh has identified a motivation for a POSA to determine the width of such gaps using image processing software and also that Kim discloses any “computer image analysis” for orthodontic treatment purposes. *See supra*, §VI.A. And unlike the ’248 patent, both Kim and

⁶ I understand that Frey is not prior art because it reflects a disclosure by the inventors of the ’248 patent, but whether Frey is prior art does not change my opinions.

Salah teach a POSA to use teeth positions, not aligner fit, to assess orthodontic treatment progress. *See supra*, §§VI.A., VI.C.

171. Dr. Foroosh next says “[t]o the extent that Salah does not expressly teach using a deep learning device trained by means of a learning base for image analysis, this would have been obvious to a POSA and is taught by Maninis.” *Id.* ¶75. He acknowledges that Maninis does not address dental images, but says that its techniques are “broadly applicable.” *Id.* ¶76. Although Salah expressly teaches that “deep learning” is “not preferred” for this purpose, Dr. Foroosh appeals to the “trends developing in computer vision” as motivating its application and “confirming that Maninis would be a natural selection for such a technique.” *Id.* But Dr. Foroosh ignores that Salah teaches that its image analysis is “best carried out ... on the patient’s own device, preferably using the [image acquisition device]” (*i.e.*, a cellphone), which would have discouraged a POSA from using a resource intensive methodology, like deep learning. *See supra*, §VI.C. He further ignores that Salah teaches away from using low contrast contours, as would be presented by an aligner on the patient’s teeth, calling them “unreliable” and teaching that they should be filtered out. *Id.* Nonetheless, Dr. Foroosh opines that “a POSA would have been motivated to adapt Salah and Kim’s image acquisition and analysis devices to employ a deep learning architecture as taught by Maninis.” *Id.* ¶78.

172. Apparently recognizing that Salah discloses only generic contour detection, Dr. Foroosh relies on Maninis as “providing good reasons to employ deep learning techniques for **object** and contour detection.” *Id.* ¶77. Dr. Foroosh suggests that “Maninis’s deep learning architecture would provide specific functionalities required by Salah, Kim, and Invisalign, including **the ability to identify objects** (such as the teeth and aligners depicted in each image) and to detect contours (such as the edges of teeth and aligners useful to assess aligner fit as taught by Invisalign).” *Id.* ¶78; *see also id.* ¶98 (noting that “Maninis’s ‘deep learning architecture’ is illustrated in Fig. 2” of Maninis). Dr. Foroosh thus proposes that “a POSA would have been motivated to incorporate a trained neural network as taught by Maninis to detect edges **and identify objects in an image to help assess aligner fit based on separation between teeth and aligner[.]**” *Id.* ¶80.

173. Dr. Foroosh describes the specifics of his proposed application of Maninis when addressing Step (b”) of ’248 patent claim 1. *Id.* ¶¶93-109. As Dr. Foroosh demonstrates by reference to Salah Figure 6b, the contours in Salah are not “semantic” and do not identify edges as associated with distinct objects. *Id.* ¶¶95-96; EX1008, 34:5-14, Fig. 6b. Dr. Foroosh thus points to Maninis as teaching “methods of detecting edges (or ‘contours’) in images using a deep learning device **in order to perform image classification**, using a process referred to as “Convolutional Oriented Boundaries (COB),” which is a neural network (herein,

“Maninis COB neural network”). EX1002, ¶¶97-99.⁷ Dr. Foroosh confirms that his proposed combination depends on application of the Maninis COB neural network

⁷ Dr. Foroosh indicates that “the specific types of neural networks overlap between the ’248 patent and Maninis” because “the ’248 patent lists ‘Microsoft ResNet’ and ‘VGGNet’ as two examples of neural networks, matching Maninis’s presentation of ‘results both using VGGNet and ResNet.’” EX1002, ¶99. ResNet and VGGNet are image classification networks. When used intact, as described in the ’248 patent, they would perform what the ’248 patent refers to as “global analysis,” producing image attributes. Neither ResNet nor VGGNet produce contours of any kind in their “native” form. Maninis incorporates a portion of these networks into the Maninis COB neural network (*i.e.*, as a “Base CNN”). EX1007, 2-3. Specifically, in Maninis, the classification portion (final layers) of these networks is removed, and the pretrained initial layers are used as feature detectors to improve the Maninis COB neural network’s determination of generic contours (not semantic contours or object identification). Maninis never proposes that either network is used intact on its own, nor does Dr. Foroosh propose that these networks would be used in his combination. It is unsurprising that both the ’248 patent and Maninis refer to these well-known neural networks, and this overlap has no apparent connection to and provides no

when addressing claim 13, which depends from claim 1 and further requires that “in step b”), the deep learning device is a neural network.” In addressing claim 13, Dr. Foroosh cites *only* the Maninis COB neural network, reproducing the architecture of the Maninis COB neural network (and the subsequent contour analysis) depicted in Maninis Figure 1. EX1002, ¶146-48 (citing EX1007, Fig. 1 (“Overview of COB”) and Fig. 2 (“Our deep learning architecture”)).

174. Dr. Foroosh proposes that the Maninis COB neural network would be trained on “localized pixel-wise semantic annotations,” which are “indications, pixel-by-pixel, of the locations of different objects in each image, identifying what object is depicted in each part of the image.” *Id.*, ¶100. “Maninis’s deep learning algorithm uses the contours it generates to define boundaries around regions and to identify the objects in each region of the image. ... Objects are identified in regions by ‘label[ing] each of the detected pixels with the corresponding semantic class.’” *Id.*, ¶101 (citing EX1007, Abstract, 11-12). According to Dr. Foroosh, this would produce “values assigned as labels” to each pixel for the set of pixels forming the boundary of the detected object. EX1002, ¶101.

support for Dr. Foroosh’s proposed combination, nor does it impact my own opinions regarding the same.

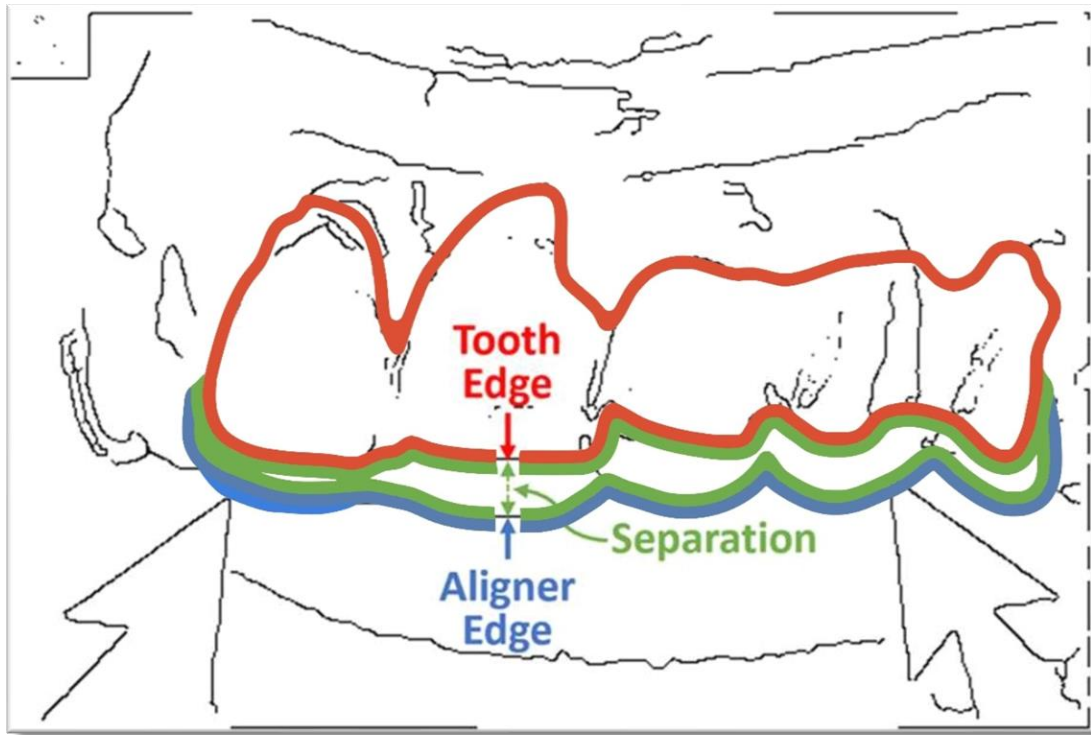
175. Dr. Foroosh further proposes that, in his combination, a POSA would have been motivated to train the Maninis COB neural network to “identify teeth and aligners depicted in orthodontic images by detecting their contours and identifying the objects found within each contour.” *Id.*, ¶102. The training images in the “learning base” would thus, in Dr. Foroosh’s proposal, be “images of dental arches of patients wearing aligners in a service position ... taken in the context of Kim and Salah’s cellphone imaging techniques” and “would be labeled with appropriate annotations,” including pixel-wise annotations of each tooth, the aligner, and the “gaps” or “spaces between them.” *Id.*, ¶¶102-103. Similarly, Dr. Foroosh proposes that a “POSA would then have set the output layers of Maninis’s neural network to provide the required object identifications corresponding to the labels used to identify individual teeth, aligner portions, and spaces between them[.]”⁸ *Id.*

⁸ In his ’409 patent IPR deposition, Dr. Foroosh confirmed that, in his opinion, the Maninis COB neural network “is providing category-specific contours” and that “while it’s producing the contours or boundaries, it is also identifying those contours or boundaries as associated with a particular object,” even when used alone as described in Maninis Section 5 regarding “low-level applications.” EX2048, 211:9-212:15; EX1007, 5-9. Dr. Foroosh likewise suggested that the Maninis COB neural network is capable of object detection and that Maninis discloses that its COB neural

176. I have illustrated below an example of the labeling and output that Dr. Foroosh proposes in his combination using the image Dr. Foroosh cites from Invisalign.⁹ In the below, the pixels corresponding to the contours of teeth (red), the aligner (blue), and the space between (green) have been annotated to identify those objects. While in the below, this is represented by different colors, what matters is that each pixel is assigned a distinct value corresponding to the object it represents.

network is only “evaluat[ed] against” Fast-RCNN. *Id.*, 215:7-216:22; EX1007, 9-13. Both of these statements are incorrect. *See* §VI.D.

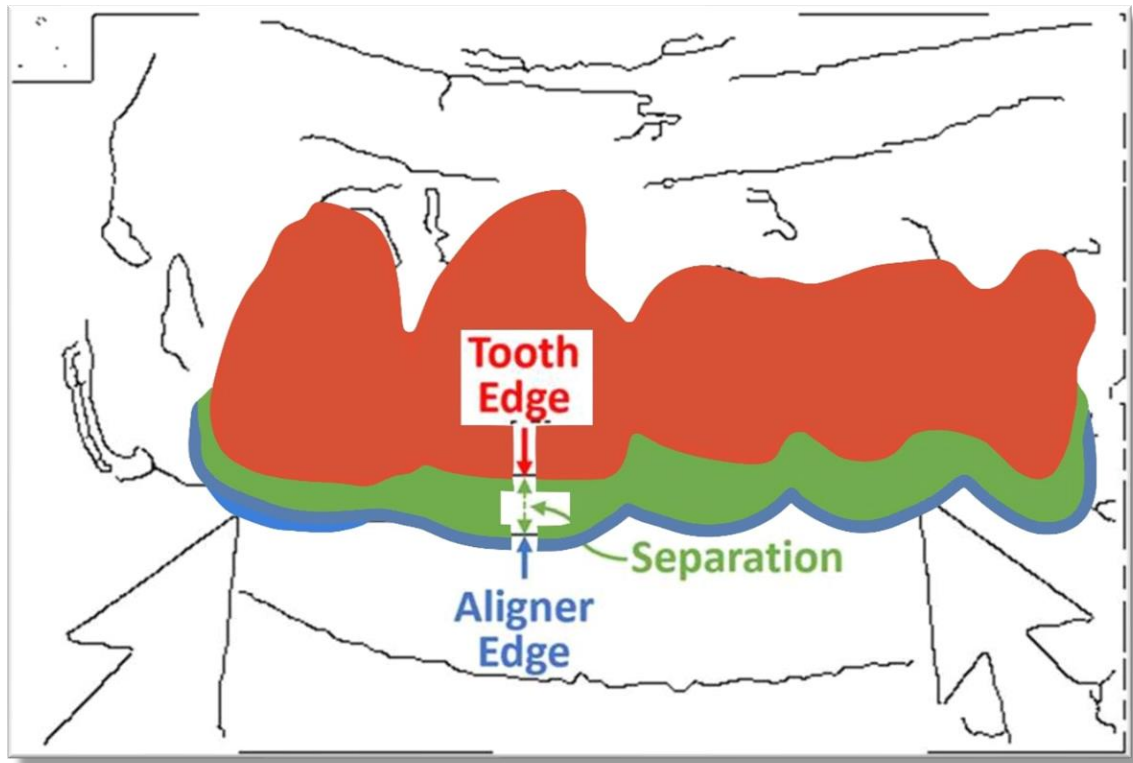
⁹ As discussed in Section VI.B., I understand from Dr. Kusnoto that this image is inapt for illustrating an aligner in a service position, but like Dr. Foroosh, I will use it herein as an example because of the exaggerated visibility of the separation between teeth and aligner.



In Dr. Foroosh's proposal, each training image is labeled in this manner, and when the trained neural network analyzes a new image, it produces semantic boundaries similar to the above training image, with boundary pixels identified and classified as corresponding to an object the neural network has determined as present in a region of the image.

177. Dr. Foroosh also cites and refers to Maninis's discussion of semantic **segmentation**, which is similar to semantic boundaries except that all pixels identified as corresponding to an object are labeled rather than only the boundary pixels (*i.e.*, the pixels inside the boundaries are also labeled). A POSA would have understood that it was possible to translate one to the other, and the distinction does not impact my opinion's regarding Dr. Foroosh's proposed combination. Below is

a visual example of a semantic segmentation based on the example of semantic contours above:



B. Dr. Foroosh's Proposed Combination is Inoperable

178. As described above, Dr. Foroosh's proposed combination requires that the Maninis COB neural network is trained on a learning base comprising historical images with semantic contours (*i.e.*, the contours of individual objects and the objects they represent), and analyzes images to determine the same. *Supra* §VII.A. Dr. Foroosh appears to fundamentally misunderstand the Maninis COB neural network. As also described above, the Maninis COB neural network is incapable of such training on historical images with semantic contours or analysis to determine semantic contours. *Supra* §VI.D. Because this functionality is critical to Dr.

Foroosh's proposed combination for each challenged claim, a POSA would have recognized that Dr. Foroosh's proposed combination is inoperable.

179. A POSA would have understood that the Maninis COB neural network is capable of being trained *only* on contours, and is capable of identifying in an image *only* contours. *Supra* §VI.D. Those contours are aggregated into hierarchies ("UCM") to identify regions ("segments") likely depicting objects, but the Maninis COB neural network never determines *what those objects might be*. It cannot. By way of an example from Maninis, while the Maninis COB neural network can detect *the contour* of a cat in an image (*see* EX1007, Fig.9, Row 3 middle image), it has no way to determine that what it contoured *is* a cat (as opposed to a train, a person, a motorcycle, or anything else). As Maninis itself expressly discloses, another neural network is required to make that kind of "semantic" determination.¹⁰ *Id.*, 9-13.

180. Accordingly, a POSA would not have been motivated to combine, and would have instead been discouraged from combining, Maninis with Kim, Invisalign, and/or Salah as Dr. Foroosh has proposed to determine semantic contours, and would not have had a reasonable expectation of success in doing so.

¹⁰ For this reason and the other reasons set forth in this Section, Dr. Foroosh's combination is also inoperable for identifying semantic segments, not just semantic contours.

In particular, a POSA would have understood that neural networks are not general-purpose tools: rather, they are only useful once they have been trained, and can only recognize what they have been trained to recognize. Each neural network trained for an image recognition task is unique and specific to the annotated images on which it has been trained, and its success (then and now) depends on the quality of that training. Because a POSA would have known that the Maninis COB neural network could not be trained to “identify teeth and aligners depicted in orthodontic images by detecting their contours and identifying the objects found within each contour,” a POSA would not have been motivated to combine Maninis’s COB neural network with Kim, Salah, or Invisalign to perform this task, as alleged by Dr. Foroosh. EX1002, ¶102.

181. Likewise, because a POSA would have known that the Maninis COB neural network could not be trained to identify objects, a POSA would not have reasonably expected success in using “Maninis’s neural network to provide the required object identifications corresponding to the labels used to identify individual teeth, aligner portions, and spaces between them,” as also alleged by Dr. Foroosh. *Id.*

182. Dr. Foroosh has not identified a “deep learning device, trained by means of a learning base” that could be used for image analysis, as required by all of the challenged claims, and a POSA would not have had a reasonable expectation

of successfully practicing the claimed image analysis, which requires “analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for at least one tooth attribute of an ‘analysis tooth zone’” and “an image attribute of the analysis image.” EX1001, *e.g.*, claims 1, 15.

183. Dr. Foroosh’s opinions are limited to the process referred to as “Convolutional Orientated Boundaries (COB)” using the Maninis neural network architecture illustrated in Figure 2. EX1002, ¶¶96-98; *supra* § 172. Dr. Foroosh has offered no opinions nor evidence regarding the motivations and expectations of a POSA with respect to any combination that requires the use of *both* the Maninis COB neural network *and* a second neural network, nor has Dr. Foroosh discussed or cited any evidence regarding the operation of the other neural networks in Maninis. Dr. Foroosh has likewise offered no opinions regarding the motivations of a POSA to use the Maninis COB neural network to determine only generic contours, as opposed to his proposed semantic contours. As discussed further below, a POSA would have had no such motivation regarding any of these undisclosed theories. *Infra* §§VIII.D-H.

C. Kim, Invisalign, and Salah Fail to Disclose Or Render Obvious a Method of Assessing Aligner Fit Using An Aligner Image, and there is No Motivation To Combine the References as Proposed

184. The challenged claims recite a method of assessing aligner fit using an image of an aligner in a service position, including captured on a cellphone.¹¹ As an initial matter, none of Dr. Foroosh's references individually disclose this technique: Kim and Salah concern assessing tooth positions and do not disclose images taken of an aligner in a service position, and Invisalign similarly does not disclose assessing aligner fit from an aligner image taken with a cellphone. *See supra*, §§VI.A-C. Even taken together, Dr. Foroosh's combination of references fails to teach or render obvious acquisition of aligner images for use in a method of assessing aligner fit.

¹¹ While claim 15 of the '248 patent does not require acquisition of an aligner image using a cellphone, Dr. Foroosh does not suggest that a POSA would have been motivated to use any other acquisition device to capture an image. *E.g.*, EX1002, ¶¶ 153-158 ("The omission of the requirement of a cellphone, however, merely means that the use of a cellphone is optional in claim 15. Thus, my analysis above [regarding elements of claim 1] applies equally ..."). Accordingly, my opinions apply equally to all challenged claims of the '248 patent.

185. First, Dr. Foroosh’s motivation to combine Invisalign and Kim is premised on the incorrect assertion that Kim teaches the use of photographs of seated aligners in remote treatment. EX1002, ¶71 (“Invisalign corroborates Kim’s teachings regarding imaging an aligner in a service position on teeth, showing how such an aligner may be assessed to determine if it fits appropriately.”); ¶72 (“In view of Kim’s disclosure of orthodontic checking of a photo or video of an aligner currently worn by the patient and Invisalign’s teachings of how to evaluate aligner fit, illustrated by an image of an aligner worn on a patient’s teeth, a POSA would have had reason to analyze images of the teeth of a patient wearing an aligner in a service position to evaluate whether the aligner fits appropriately.”). As explained above, however, Kim is concerned with assessing teeth alignment, not the distinct process of assessing aligner fit. To assess the alignment of teeth, Kim teaches the acquisition of images of the patient’s teeth without aligners.

186. The “clear aligner information transmission program” of Kim relates to photographs of aligners *on their own*. *Supra* §VI.A; EX1005, [0066], Figs. 1 and 6. There is no suggestion in Kim that these photographs of aligners should be used for diagnostic or treatment purposes; rather, Kim instructs that the photographs of patients’ teeth that comprise the separate teeth alignment information report should be used for those purposes. EX1005, [0085] (“the orthodontist may check the transmitted teeth alignment report information per patient using the orthodontist

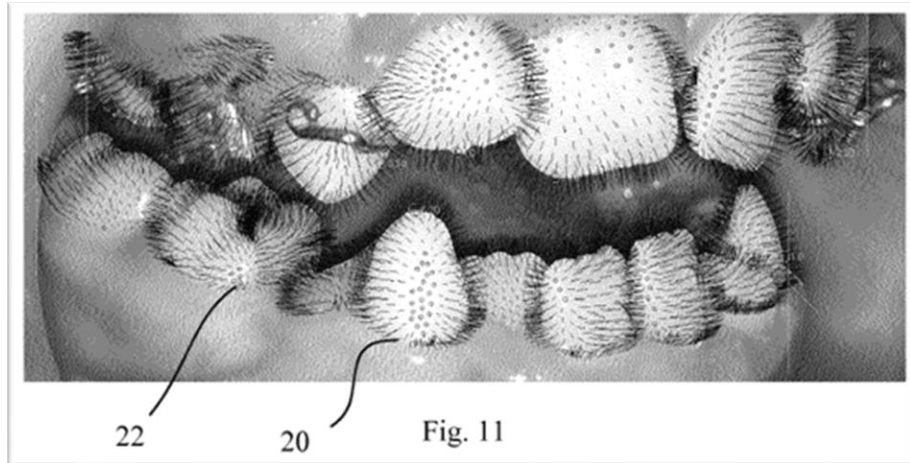
terminal”), [0087] (“the orthodontist may remotely provide accurate diagnosis by checking [the various photos] capable of showing the teeth alignment state of the orthodontic patient”); *see also id.*, [0063], [0093]. Thus, the methods of Kim do not disclose the use (or usefulness) of photographs showing seated aligners or suggest that such photographs would be useful to an orthodontists’ remote or virtual evaluation of a patient’s treatment.

187. Second, Dr. Foroosh is incorrect that “Invisalign confirms that analyzing an image of an aligner in a service position on the patient’s teeth (such as a photo or video taken by a cellphone, as taught by Kim) can produce measurements useful for orthodontic assessment.” EX1002, ¶88. Invisalign discusses the collection of patient images only for the purposes of aligner design and manufacture, *i.e.*, before the aligners exist, and there is nothing in Invisalign that indicates that such images could be taken with a cellphone. Rather, a POSA would have understood that such dental photographs would have been taken with a more sophisticated camera, such as a DSLR, and in-office with bright and controlled lighting by a trained dental professional. I understand from Dr. Kusnoto that, in 2017, in-office photographs taken by a trained dental professional with a professional-grade camera were of significantly higher quality than patient-created photographs using a cellphone would have been.

188. Further, while Invisalign notes that a separation of greater than 1 mm between the tooth and the aligner is indicative of a fit issue, Invisalign nowhere suggests that aligner fit can be assessed simply by looking at an aligner image. Instead, Invisalign sets out various steps that should be performed in person to evaluate aligner fit. EX1006, ii (“At each appointment ... [e]valuate for tight contacts with unwaxed floss and relieve with finishing strips if present.”). Indeed, I understand from Dr. Kusnoto that, to assess aligner fit, an orthodontist would visually and manually examine the aligner, using their hands to check for rocking (movement of one part of the aligner off of a patient’s teeth when pressure is applied). I further understand from Dr. Kusnoto that, when an aligner is properly seated, even if there are fit issues, they are difficult to diagnose solely by visual inspection by a dental professional. For example, I understand from Dr. Kusnoto that even a trained orthodontist cannot reliably tell if an aligner is within 1 mm of the tooth without further manipulation, such as using a tooth scaler or other tool to evaluate fit and provide scale. EX1006, 1 (Invisalign describing the use of a scaler to “check retention”). Moreover, based on my review of Invisalign as confirmed by Dr. Kusnoto, Invisalign does not disclose any methods for measuring whether an aligner is within 1 mm of the tooth, apart from visual approximation. A POSA would have understood that this is because, in 2017, orthodontists did not quantitatively assess aligner fit and did not measure the distance between an aligner and the teeth.

189. For these reasons, upon review of Invisalign, a POSA would not have been motivated to modify the approach in Kim so as to measure the distance between tooth and aligner from a cellphone image captured by an untrained professional. Nothing in the combination of Kim and Invisalign suggests that the in-person approximation, combined with in-person visual and manual inspection and manipulation of the aligner, could be successfully replaced with computer-implemented analyses of patient-initiated photographs captured using a cellphone, or would have motivated a POSA to do so.

190. Combining Salah with Invisalign and/or Kim also would not have motivated a POSA to acquire images showing an aligner in a service position for use in a method of assessing aligner fit. While Salah includes one image that includes a device, I understand from Dr. Kusnoto that a POSA would have recognized that Figure 11 shows a bonded expander on the upper teeth that could not be removed by the patient prior to taking the photograph—and not a clear aligner, which could have been readily removed before taking the picture. EX1008, Fig. 11 (shown below).



191. Further, Salah assesses tooth positions, it does not provide any methodology to measure aligner fit or suggest that it could be assessed from an image. To the contrary, Salah teaches that “[i]n the case of a contour, for example, contrast analysis provides more or less reliable information: an area of high contrast can be assimilated to an area corresponding to a contour with a high probability, and the quality of the points of this area will be high. On the other hand, a low-contrast area, such as a blurred zone, can be assimilated to an area corresponding to a contour with a low probability, and the quality of the points of this area will be low.” EX1008, 60:17-22. Salah further recommends filtering and retaining only the outlines identified with a high probability, calling those identified with a low probability “unreliable”. EX1008, 61:2-4. In other words, Salah teaches that high contrast is *required* to achieve a quality analysis. I further understand from Dr. Kusnoto, that a clear aligner does not present a high contrast with teeth or the

mouth. Thus, Salah teaches away from using contour detection with images involving a clear aligner.

192. Finally, Dr. Foroosh opines that Frey “confirms that acquiring digital photographs of patient’s teeth, including patient-acquired cellphone images of teeth of patients wearing aligners in a service position, was known to be a useful practice for assessment of a patient’s orthodontic treatment status.” EX1002, ¶90. I understand from Dr. Kusnoto that this is not correct. It was rare to use dental images captured by patients in 2017 and there would have been many challenges with acquiring images in this way for the purposes of assessing aligner fit. Consistent with that understanding, Frey only suggests obtaining a video that the orthodontist would then remotely visually inspect.

193. I note that Dr. Frey published another post on his blog approximately two weeks earlier evaluating Dental Monitoring’s “GoLive” product, which was “designed for Invisalign patients” and which involved taking “photos ... with aligners in place” to advise the patient whether to “advance to the next aligner.” EX2053, 6. Dr. Frey states that “DM GoLive is an interesting concept, but weekly photo submissions add additional compliance burdens on patients and in my view would be counterproductive to treatment,” further noting that “[t]he patients compliant enough to complete these weekly evaluations aren’t the ones need[ing] a service to help titrate their aligner wear effectively.” *Id.*, 7. Dr. Frey also noted the

“additional costs” as a discouraging factor. Thus, contrary to Dr. Foroosh’s suggestion, Dr. Frey’s blog confirms that there was no motivation to acquire images of a patient’s dental arch with aligners in a service position to assess aligner images, and indeed, Frey teaches away from doing so.

194. In conclusion, claims 1 and 15 of the ’248 patent recite “[a] method for assessing the shape of an orthodontic aligner” that requires “acquisition of at least one image at least partially representing the aligner in a service position.” None of Dr. Foroosh’s references taken alone or together disclose or render obvious the acquisition of aligner images for use in a method of assessing aligner fit.

D. Kim, Invisalign, Salah, and Maninis Fail to Disclose Or Render Obvious the Claimed Analysis of an Analysis Image By Means of A Deep Learning Device Trained by a Learning Base, and there is No Motivation To Combine the References as Proposed

1. A POSA Would Not Have Been Motivated To Combine Maninis with Kim, Invisalign, and Salah

195. A POSA would not have been motivated to combine Maninis with Kim, Invisalign, and Salah to arrive at the claimed analysis of an aligner image.

196. As an initial matter, Dr. Foroosh does not explain how or why a POSA would have been aware of or located Maninis. In my opinion, a POSA would not have known of Maninis. Maninis is an academic paper and makes no mention of the use of its neural network architecture for analysis of dental or orthodontic images.

197. I further disagree with Dr. Foroosh that Salah would have “motivate[d] a POSA to apply Maninis’s teachings to the imaging apparatus of Kim and Salah.” EX1002, ¶76. On the contrary, Salah is the only cited prior art reference that even refers to “deep learning” in the context of dental image analysis, and Salah teaches that “deep learning” is “not preferred.” *Supra* §VI.C. Additionally, in view of Salah’s teaching that the **analysis itself** be performed on a cellphone (EX1008, 33:6-8, 13:24-28, 16:14-21), Salah actually discourages using deep learning, which is highly resource intensive and thus would have been less suitable for use on a cellphone in 2017. Indeed, even though Salah suggests that “deep learning” “can ... be used” to determine **discriminant information**, it would not have taught a POSA that deep learning would be a “useful option” for determining contours when Salah’s methods could be achieved through a less resource-intensive technique. *See supra* §VI.C. Salah further teaches away from using any method that would rely on low contrast contours, noting that they are “unreliable” and preferably filtered out. *Id.*

198. Furthermore, Salah’s disclosure of “deep learning” is exclusively in the context of updating a 3D model. *Id.* Dr. Foroosh attempts to extrapolate the disclosure of Salah to apply to all computer image analysis of dental images. EX1002, ¶74. Dr. Foroosh ignores, however, that the technique in Salah requires a detailed, 3D scan and/or model of the patient’s teeth against which subsequent updated images can be compared. *See supra* §VI.C. Only once the discriminant

information detected in the updated images are vetted against a corresponding virtual image of the 3D model, and the variation in perspective and quality of the image has been evaluated and accounted for, can useful information regarding the positioning of the patient's teeth be determined. *Id.* Nowhere does Dr. Foroosh propose that a POSA would have been motivated to create 3D models of each patient's teeth, let alone model aligners in a service position. And nowhere does Salah suggest contour detection of an aligner image for the purposes of analyzing aligner fit. A POSA would have had no reason to view Salah's disclosure as a general motivation to apply "computer image analysis to dental images for orthodontic treatment purposes," let alone to images of aligners using a "not preferred" technique.

199. Nothing in Salah would have taught, let alone motivated, a POSA to determine *generic* contour information using deep learning. Generic contours would not be useful in the combination Dr. Foroosh proposes here. That is because simply determining contours, without knowing whether they correspond to a tooth, aligner, gum, tongue etc. would not enable a POSA to assess aligner fit. Thus there would have been no perceived benefit of performing this resource intensive analysis. Similarly, nothing in Salah would have taught, let alone motivated, a POSA to determine *semantic* contour information using deep learning. Determining semantic contours is not necessary for the purposes of Salah and Salah discloses only the detection of generic contours.

200. Dr. Foroosh further opines that “[a] POSA would view Maninis as providing good reasons to employ deep learning techniques for object and contour detection.” EX1002, ¶77. Dr. Foroosh specifically identifies that Maninis’s COB technique “is compatible with different CNNs,” “is applicable in a ‘modular’ fashion,” and ““can incorporate and benefit from future improvements in the base CNN.”” *Id.* (citing EX1007, 2). Here, Dr. Foroosh’s reasoning is circular because it assumes that a POSA would have already decided to use a neural network. Additionally, Dr. Foroosh does not explain how or why the alleged benefits would have motivated a POSA to include Maninis in the combination of Kim, Invisalign, and Salah, and I disagree that they would. The fact that Maninis can be applied to many different contexts is irrelevant to whether it would be desirable to incorporate it into this particular combination. That Maninis’s COB neural network might “benefit from *future* improvements” is of minimal and speculative value, particularly in view of the significant costs, downsides, and compatibility concerns addressed further below.

201. Dr. Foroosh also states that Maninis describes that its technique offers “improvements” over “recent approaches to contour detection using deep learning,” specifically alleging that “Maninis demonstrates substantially faster object edge and contour detection and image classification relative to state-of-the-art techniques” and describes improvements in semantic segmentation. *Id.*, ¶77. As described

above, Dr. Foroosh misunderstands Maninis’s disclosures. The neural network on which Dr. Foroosh relies—Maninis’s COB neural network—is **only** able to determine generic contours. *Supra* §VI.D. It cannot perform the task that Dr. Foroosh alleges a POSA would have been motivated to use Maninis’s COB neural network to conduct, namely semantic segmentation. *Id.* The reported speed on which Dr. Foroosh relies in Maninis (0.79 seconds) reflects the identification of *generic* contours—not the objects they define. *Id.* And as noted above, a POSA would not have been motivated to detect only generic contours as they would not have been useful in assessing aligner fit from an aligner image (assuming a POSA would have been motivated to attempt to do so), according to Dr. Foroosh’s proposed combination.

202. Maninis does not report any speed for object detection and as such, it cannot demonstrate “substantially faster ... image classification relative to state-of-the-art techniques” as alleged by Dr. Foroosh. *See id*; *see also* EX1002, ¶77. Rather, as I explained above, the authors of Maninis determined semantic contours only by using a separate, prior art neural network in parallel with their COB neural network. *Supra* §VI.D. Dr. Foroosh has never suggested that a POSA would have been motivated to use more than one neural network. *E.g.*, EX1002, ¶78 (“Maninis’s flexibility and prospect to improve performance would motivate a POSA to analyze images using **a** deep learning device, as taught by Maninis”); ¶97 (“Maninis teaches

methods of detecting edges ... in images using **a** deep learning device ... using **a** process referred to as ‘Convolutional Oriented Boundaries (COB).’”); ¶102 (“a POSA would have trained **a** neural network like Maninis’s”); ¶105 (“Maninis teaches image analysis with a deep-learning device, in the form of [a] computing device with **a** trained neural network”); ¶156 (“As I discussed above for claims 1 and 8, it would have been obvious to train **a** neural network ...”). Furthermore, Dr. Foroosh has provided no information regarding the speed or efficiency of either the *second* neural network, or the *third* process of integrating the output of the Maninis COB neural network with the output of the second neural network to assign semantic classes to the boundaries. Importantly, Dr. Foroosh never suggests how the alleged speed of Maninis techniques would motivate a POSA to apply Maninis to Kim, Invisalign, and Salah; a POSA would not have cared about the speed of a technique if there was no motivation to apply that technique at all (or if the technique was inoperable for the alleged purpose).

2. A POSA Would Have Recognized the Significant Time and Expense Associated with Implementing Maninis for the Alignment Techniques of Kim, Invisalign, and Salah

203. Dr. Foroosh alleges that a POSA would have been motivated to train the Maninis COB neural networks on a learning base of images “representative of the types of images likely to be seen during ordinary use of the imaging device, such as real historical photographs taken of patient’s teeth wearing aligners with the

device in which the trained network will be deployed (*e.g.*, a cellphone as taught by Kim and Salah).” EX1002, ¶123. However, the images comprising the training sets described in the Maninis reference are standard training sets for computer vision competitions and benchmarking and do not include dental images. *Supra* §VI.D, VIII. Dr. Foroosh has not identified any such images that already existed as of July 2017. And as explained above, neither Kim nor Salah teaches a POSA to capture images showing an aligner in a service position. *Supra* §§VI.A, VI.C., VIII. Rather, both are concerned with methods of tracking tooth movement that use images without an aligner present. *Id.* Thus, it would not have been possible to create a learning base, as claimed, using the images from Kim and Salah.

204. Dr. Foroosh also does not address the time or expense associated with capturing aligner images using a cellphone, but I understand from Dr. Kusnoto that it would take significant time and cost to do so. *Supra* §VI.D.2. To the extent these images were collected from dental practitioners, I understand from Dr. Kusnoto that patient anonymization of dental and orthodontic photos would be a necessary step for compliance with privacy laws, such as HIPAA, and a POSA would have known that this can be prohibitively expensive. *Id.*

205. Once acquired, a POSA would also need to label these images with pixel-wise semantic boundaries identifying teeth, the aligner, and the gap between them, to be consistent with the training Dr. Foroosh proposes in his combination.

EX1002, ¶¶102, 124-125; EX2048, 100:6-12. Pixel-wise, object-specific contours of medical images are challenging and expensive labels to apply to images in a learning base because they require precise and extensive annotation and may require dental professionals to perform the annotations. *Supra* §VI.D. A POSA would have further known that the accuracy of the annotations in the historical images in the learning base would impact the accuracy of the neural network’s analysis after training. *Id.* In fact, Maninis discloses its results using hand labeled and human verified contours. EX1007, 1-6 (describing PASCAL VOC databases).

206. The cost, time, and complexity required to obtain labeled images to practice the combination of Kim, Invisalign, Salah, and Maninis as Dr. Foroosh proposes would represent a significant disadvantage that would have further discouraged a POSA from applying deep learning to determine aligner separation. *See supra* §VI.D.

3. Dr. Foroosh’s Semantic Contours Are Not “Tooth Attributes” (Claims 7 and 14)

207. Claims 7 and 14 depend from claim 1 and require that step b”) comprise additional steps regarding the creation of the learning base, the training of the deep learning device, and the deep learning device’s analysis of the analysis image. EX1001, cls. 1, 7, 14. Claims 7 and 14 require that each “historical image” in the learning base comprise “historical tooth zones” (*i.e.*, “one or more zones each representing a tooth”) “to each of which ... a tooth attribute value is assigned.”

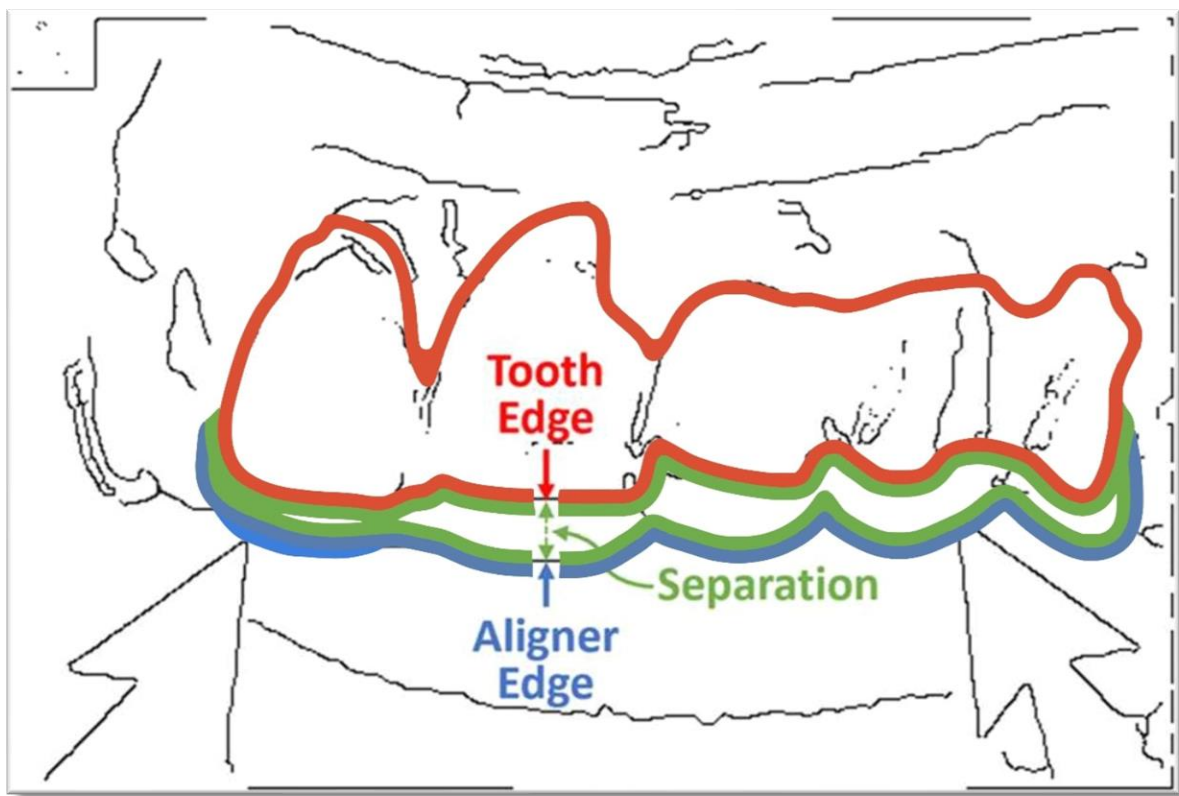
Claims 7 and 14 further require that, after training the deep learning device on a learning base comprising “more than 1000” such “historical images,” “the deep learning device ... determine at least one probability relating to: the presence, in a location of said analysis image, of an analysis tooth zone; and the attribute value of the tooth represented on said analysis tooth zone.” *Id.*

208. Consistent with the specification, claims 7 and 14 thus distinguish between a “tooth zone,” which corresponds to the “zone” or “location” of a tooth in the analysis image, and a “tooth attribute,” which is an “attribute” of the tooth represented “*on said ... tooth zone.*” In other words, the “tooth zone” defines where the tooth is, and the “tooth attribute” describes the tooth in that zone. Accordingly, the claim requires that the deep learning device determine a probability related to not only “*the presence, in a location*” of the tooth zone, but also to the “*value*” of the corresponding tooth attribute. In both claims 7 and 14, the tooth attribute labeled on the historical images must “relat[e] to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented.” *Id.*

209. Dr. Foroosh alleges the following regarding how his proposed combination satisfies the “tooth attribute value” limitation in claims 7 and 14: “With labels to indicate the pixels corresponding to teeth, the pixels corresponding to aligners, and the pixels corresponding to the gap between the two, each historical image would include tooth attribute values assigned directly to the image by the

labels. For example, the boundaries between regions would directly indicate the edges of teeth, or of the aligner, both of which are tooth attributes Further, labeling the pixels between each tooth and the corresponding aligner surface (*i.e.*, the gap), the image would include a direct label for the tooth of the separation between the tooth and the aligner[.]” EX1002, ¶125. I disagree that any of the foregoing satisfies claims 7 and 14.

210. Consistent with his description of the proposed combination throughout his report, Dr. Foroosh proposes that the “historical images” are labeled on a pixel-wise basis, with each pixel identified as “corresponding to teeth,” “corresponding to aligners,” or “corresponding to the gap between the two.” *Id.* Using Dr. Foroosh’s own example contours, below is an approximate illustration of the semantic contours Dr. Foroosh proposes, with the pixels identified as teeth contoured in red, the pixels identified as aligners contoured in blue, and the gap between the two contoured in green:



211. As noted above, claims 7 and 14 require that the “tooth attribute” is an attribute of the tooth “on said ... tooth zone.” Thus, merely identifying which pixels comprise the tooth zone, or which pixels correspond to adjacent objects, is not assigning an attribute value to any historical tooth zone. Indeed, Dr. Foroosh’s proposal contrasts with tooth attributes identified in the patent, which include a “maximal” or “mean separation” along the edge of a specific tooth. EX1001, 28:50-55. Dr. Foroosh does not propose that any tooth zone is labeled with the value of *the presence or amplitude* of a related separation; he only proposes that pixels representing a gap or an aligner are identified as such. An indication that a pixel depicts “separation” or “aligner” is not an attribute of a tooth represented by a

historical tooth zone. While one could analyze the separation between the labeled boundaries and determine a value for an attribute of a tooth in an adjacent tooth zone (e.g., magnitude of separation from the aligner), Dr. Foroosh does not propose that any such value is present in the “historical images,” and for good reason. As explained below (§VIII.D.4) there would be no way for the neural networks Maninis discloses as determining semantic segmentations to accept as training inputs or output such values in addition to the semantics already identified for each pixel (e.g., “tooth,” “gap,” and “aligner”).

212. Semantic boundaries and edges (and segments) identify the *location* in an image of an object, not its attribute. Dr. Foroosh’s proposal that the boundaries/edges are “attributes” ignores that the claims already have a term for the region identifying the location of an object: “zones.” Dr. Foroosh himself identifies the edges/contours of teeth as “analysis tooth zones,” arguing that Maninis’s “contour detection process” involves determining the presence of an “analysis tooth zone” “because ***the set of tooth edges define [an] analysis tooth zone*** that has been determined to contain a particular tooth.” EX1002, ¶132. Thus, not only are the edges of an *aligner* not a “tooth attribute,” they are not an “attribute” at all. Likewise, the boundary of a tooth defines the “tooth zone,” which is distinct from a “tooth attribute” *in that zone*. While, according to Dr. Foroosh’s proposed combination, the pixels in the “historical tooth zone” would be “labeled ... as

corresponding to teeth,” that too is indistinct from the “historical tooth zone” itself, which is a “zone[] ... representing a tooth.” Labeling the pixels in the tooth zone as pixels depicting teeth only establishes that zone as a “tooth zone” but does not constitute a value of an “*attribute*” of the tooth depicted therein.

213. Furthermore, boundaries and edges (and segments) are not “values” as required by the claim. Consistent with the claim language, boundaries and edges—which are most analogous to “zones”—indicate the “presence, at a location” of an object. The claims require that the “historical images” are labeled with a “*value*” for a tooth attribute. Even if the boundaries and edges of the teeth and/or aligners were somehow “tooth attributes,” Dr. Foroosh cannot identify a “value” that could be labeled in the historical image that would represent boundaries or edges, each of which are represented as *images* (i.e., bitmaps of individual pixel values).¹²

¹² Dr. Foroosh suggests in a parenthetical that the “analysis tooth zone” may be drawn such that it is a “region containing a tooth and an aligner surface separated by a gap” rather than the tooth alone. EX1002, ¶ 106. This embodiment does not affect the above analysis – neither the “zone” itself nor the labeling of pixels as corresponding to the aligner or gap are “tooth attributes,” nor are their edges “values,” regardless of whether the aligner and gap are included within or adjacent to the analysis tooth zone.

214. For these reasons, “tooth and aligner edges” are not tooth attributes or tooth attribute values of the historical images or the analysis images, as required by claim elements 7(1), 7(3), 14(1), and 14(3). And Dr. Foroosh has failed to identify a neural network that could be trained on values corresponding to the “size of the gap” in addition to the semantics already identified for each pixel (*e.g.*, “tooth,” “gap,” and “aligner”). Accordingly, a POSA would not have been motivated to and would not have had a reasonable expectation of successfully practicing claims 7 or 14 as alleged by Dr. Foroosh.¹³

4. Dr. Foroosh’s Semantic Contours Are Not “Image Attributes” (Claim 8)

215. Claim 8 depends from claim 1 and requires that step b”) comprise additional steps regarding the creation of the learning base, the training of the deep learning device, and the deep learning device’s analysis of the analysis image. EX1001, cls. 1, 8. Claim 8 requires that each “historical image” in the learning base comprise an “image attribute value,” “the image attribute relating to a separation between the tooth or teeth and the aligner represented on the historical image,” and

¹³ The foregoing applies equally to semantic contours and semantic segmentations, both of which identify the location of objects on a pixel-by-pixel basis, as discussed above.

that “at least one probability relating to **said** image attribute value” is determined during analysis. *Id.*, cl. 8. Accordingly, the same “image attribute value” must be *both* labeled on the historical images in the learning base *and* related to a probability determined by the deep learning device for the analysis image. Thus, in contrast to claims 7 and 14, which embody the “detailed analysis approach” described in the ’248 patent, claim 8 embodies the “global analysis” approach. *Supra* §58 (contrasting “**detailed analysis**” approach, wherein the deep learning device is trained on and determines “tooth zones” and corresponding “**tooth attribute** values” with the “**global analysis**” approach, wherein the deep learning device is trained on and determines “**image attribute** values”).

216. Dr. Foroosh appears to identify the following possible “image attributes” in the historical images when addressing claim 8 limitation 8(1’):

- “**‘presence’ of an aligner**” as shown by annotation of pixels as corresponding to an aligner (EX1002, ¶140)
- “**presence of a separation**” as shown by annotation of pixels as corresponding to a gap (*Id.*)
- “**the locations of each tooth and each corresponding aligner part**” (*Id.*)
- “Considered collectively, **the set of annotations** would relate to an image attribute value, as they would relate to the value of whether the aligner in the

image was well or poorly fitted, under the standard taught by Invisalign (e.g., they would depict gaps greater or less than 1mm).” (*Id.*)

- “The gap between aligner and tooth directly represents the separation between each tooth and the aligner, so **the annotation of such gaps** would also indicate an attribute value for the image[.]” (*Id.*)

217. Dr. Foroosh appears to identify the following possible “image attributes” as related to a probability determined by the deep learning device when addressing claim 8 limitation 8(3’):

- Dr. Foroosh refers generally to “**the image attributes discussed above** for step 8(1’).” (EX1002, ¶144)
- “**the distance between the teeth’s edges and the corresponding edges of the aligner**, as reflected in the boundaries of the respective tooth and aligner regions as well as in the annotation of the gap between teeth and aligner.” (*Id.*)
- “**the separations of all teeth from the aligner**,” apparently a “list” thereof (*Id.*)
- “**a list of locations of all teeth edges**” (*Id.*)
- “**a list of locations of all aligner edges**” (*Id.*)

218. Dr. Foroosh ignores that the claim 8 requires that the image attribute related to the probability determined by the deep learning device must be **the same**

image attribute labeled in the historical images. Thus, each supposed “image attribute” Dr. Foroosh identifies in Step 8(3’) that he does not identify in Step 8(1’) fails to satisfy this limitation of claim 8, including “the distance between the teeth’s edges and the corresponding edges of the aligner,” “the separations of all teeth from the aligner,” “a list of locations of all teeth edges,” and “a list of locations of all aligner edges.”

219. Turning to the alleged “image attributes” that Dr. Foroosh identifies in Step 8(1’), which he identifies as also in Step 8(3’) by reference (“the image attributes discussed above”), neither the “presence of an aligner” nor the “presence of a separation” can be image attributes according to Dr. Foroosh’s own proposed combination. As described repeatedly herein, Dr. Foroosh’s proposed combination only provides for the pixel-wise labeling of an image in semantic boundaries (or segmentations). While Dr. Foroosh proposes that pixels corresponding to an aligner or a separation/gap would be labeled, that is not equivalent to labeling a “historical image” as “having an aligner” or “having a separation.” The labeling of pixels as corresponding to a certain object (*e.g.*, the presence of an aligner or gap) is a regional attribute akin to a tooth attribute, not an image attribute. The ’248 patent expressly distinguishes between regional attributes and image attributes. For example, while the patent gives the example of labeling individual teeth with teeth numbers as an example of “tooth attributes,” only a collective “list” of such tooth numbers is

described as an “image attribute.” EX1001, 26:3-14. Similarly, the patent states that “[a] tooth attribute value may be used to define a value for an image attribute,” such as “if a value of a tooth attribute is ‘tooth decayed’, the value of an image attribute may be ‘unsatisfactory dental situation.’” *Id.*, 3:59-62. While the tooth attribute value “may be used to define” the image attribute value, the attributes and values are still distinct.

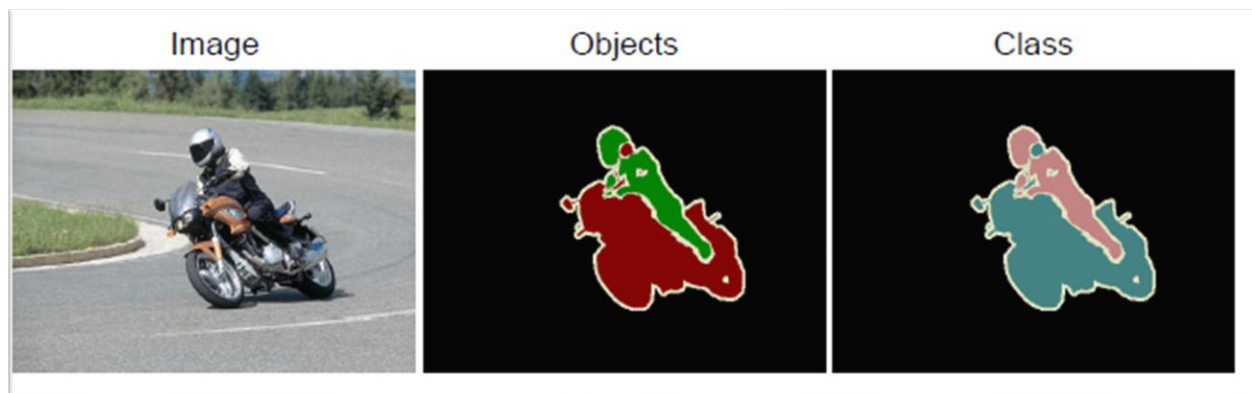
220. This is a meaningful distinction in the context of a deep learning device, where a device trained to determine probabilities that various pixels correspond to various objects produces a distinct output from a device trained to determine probabilities that the image as a whole contains those objects. The distinction drawn between those analyses in the patent (detailed analysis vs. global analysis) is reflected in the state of the art for deep learning devices, and a POSA would have recognized that they are not interchangeable. For example, some neural networks are structured and trained to have outputs relating to the entirety of an input image, as in the simple example I provide above (§V.D.) where the output of the image analysis is a probability relating to which one of the digits from 0-9 appears in the image. Other neural networks, such as the one Dr. Foroosh proposes, is trained on pixel-wise annotations and outputs values for each pixel in the analyzed image, but does not (and cannot) be trained on or output a value for the image as a whole.

221. This distinction is embodied in Maninis itself. Neither the Maninis COB neural network, which detects generic contours, nor the other neural networks with which the Maninis COB neural network is combined to produce semantic contours or semantic segmentations, are capable of being trained on images labeled with *any* “image attribute value,” including the presence of an object or a list of regional attributes. As discussed above, the labels (or annotations) on which the Mannis COB neural network is trained are bitmaps of binary values representing only a 1 or a 0 for each pixel in the training image, indicating whether or not that pixel is part of a contour. *Supra* §§VI.D., VIII. The Maninis COB neural network is not capable of accepting any inputs other than a vector of values, each value representing a pixel in an analysis image. *Id.* These vectors are not “image attributes” as they do not “relate[] to the whole image and not to part of the image” but rather, each value in the vector relates to a part of the image: a single pixel and whether it is part of a contour. *See* EX1001, 3:26-27. If a POSA added values to the training examples representing an image attribute, which would require expanding the size of the training vector, the Maninis COB neural network would have no way of mathematically accepting that information.

222. Similarly, to produce semantic contours, Maninis proposes separately running another prior art neural network that determines semantic segmentations before combining that analysis with the results of the Maninis COB neural network.

Supra §§VI.D., VIII. The training label inputs to a neural network that determines semantic segmentations is similar to the training label inputs to the Maninis COB neural network: namely, bitmap vectors with values assigned for each pixel, but rather than the values being limited to 0 or 1 and representing the boundaries of objects, the values would include other numbers, each corresponding to the class of the object to be recognized (here, teeth, aligner, gap), with the values representing the areas of the objects. *Id.* In other words, for a given object, each value associated with a pixel of the object would be set to the numeric identifier of the object's class (for instance, "teeth," "aligner," or "gap"), whereas for the Maninis COB neural network, only values associated with a pixel on the border of the object would be set to 1. *Id.* Importantly though, just as for the Maninis COB neural network, these training vector values represent detailed information about only individual parts (pixels) of the image and are therefore not "image attributes." *Id.* As with the Maninis COB neural network, such a neural network for determining semantic boundaries is incapable of being trained on a value representing anything other than values to be assigned to each pixel, including any value representing the general presence of an aligner or separation, or a list of regional attributes. *Id.*

223. Below are sample images from the training set that Maninis used to train the neural networks for semantic segmentation in Section 6.2. EX1007, 11 (referring to "PASCAL VOC Segmentation dataset"):



EX2020, 4. On the right, the value for an object or class associated with each pixel has been represented as a color for convenience. Pixels identified as being associated with background segments are colored black; pixels associated with segments representing a person are colored green/pink; and pixels associated with segments representing a motorcycle are colored red/blue. No other information—such as a value identifying the list of objects present in the image (*e.g.*, “person, motorcycle”)—can be accepted by the neural networks for training.

224. As a result, because Dr. Foroosh has proposed that the “deep learning device” is a neural network that determines contours, or semantic contours or segmentations, rather than a device that determines holistic image features, a POSA would have understood that it is impossible for those neural networks to be trained on a learning base that satisfies the limitation that “each historical image compris[es] an ... image attribute value.” *See id.* These neural networks are *incapable* of being trained on a value representing anything other than values to be assigned to each

pixel, which excludes any value that “relates to the whole image.” EX1001, 3:26-27; *see, e.g.*, EX1002, ¶125.

225. In other words, claim 8 requires the use of a deep learning device that can be trained on images that have “image attribute values,” but Dr. Foroosh’s references solely mention neural networks that can **only** be trained pixel-wise annotations. Neural networks that determine semantic contours or segmentations, as Dr. Foroosh alleges a POSA would be motivated to use, are *incapable* of accepting *any* “image attribute value.” Thus, while an image attribute value can be determined from the identified tooth attribute values *after* the image has been analyzed, in Dr. Foroosh’s proposed combination, the historical images in the learning base *cannot* comprise those same image attribute values.

226. As a result, Dr. Foroosh has failed to demonstrate that the limitation of Step 8(1’) is satisfied by the proposed combination of Kim, Invisalign, Salah, and Maninis. Furthermore, a POSA would not have been motivated to create a learning base comprising historical images labelled with image attribute values, train any of the neural networks disclosed in Maninis with such a learning base, and use such a neural network to analyze analysis images, or had a reasonable expectation of success in achieving the claimed invention in that instance because the proposed deep learning device would be incapable of being trained by means of the proposed

learning base or performing the claimed analysis, as required by claim elements 8(1) and 8(3) of claim 8.

IX. CONCLUSION

227. For the foregoing reasons, it is my opinion that Dr. Foroosh has failed to establish that any of the challenged claims would have been obvious to a POSA in July 2017 in view of Kim, Invisalign, Salah, and Maninis.

I, John Mongan, do hereby declare and state, that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, under Section 1001 of Title 18 of the United States Code.

I declare (or certify, verify, or state) under penalty of perjury that the foregoing is true and correct.

Executed July 12, 2024, in San Francisco, California.

/ John Mongan /

Dr. John Mongan