

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. IPR2023-01369

U.S. Patent No. 10,755,409

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

TABLE OF CONTENTS

I.	SCOPE OF ENGAGEMENT	5
II.	BACKGROUND	5
A.	EDUCATION AND INDUSTRY EXPERIENCE	5
B.	COMPENSATION.....	10
C.	MATERIALS AND OTHER INFORMATION CONSIDERED	10
III.	UNDERSTANDING OF THE LAW	11
A.	PERSON OF ORDINARY SKILL IN THE ART	11
B.	DATE OF INVENTION.....	12
C.	PRIOR ART	13
D.	CLAIM CONSTRUCTION	13
E.	PATENTABILITY	15
F.	OBVIOUSNESS.....	15
G.	SECONDARY CONSIDERATIONS RELEVANT TO NONOBVIOUSNESS.....	20
IV.	THE '409 PATENT	23
A.	THE EFFECTIVE FILING DATE OF THE '409 PATENT	23
B.	PERSON OF ORDINARY SKILL IN THE ART.....	24
C.	INTRODUCTION TO THE '409 PATENT	25
V.	OVERVIEW OF THE ASSERTED ART	34
A.	SALAH.....	34
B.	CARRIER.....	56
C.	MANINIS.....	71
VI.	CLAIM CONSTRUCTION.....	92
A.	“DEEP LEARNING DEVICE”	92
B.	“REAL TIME”.....	95
C.	“TOOTH SHAPE PARAMETER”	95

VII. THE '409 PATENT WOULD NOT HAVE BEEN OBVIOUS IN VIEW OF SALAH, CARRIER, AND MANINIS.....	97
A. DR. FOROOSH'S PROPOSED COMBINATION.....	99
B. A POSA WOULD HAVE RECOGNIZED THAT DR. FOROOSH'S PROPOSED COMBINATION IS INOPERABLE	103
C. A POSA WOULD NOT HAVE BEEN MOTIVATED TO COMBINE CARRIER'S ALIGNMENT TECHNIQUE WITH SALAH'S IMAGE ACQUISITION METHODS AND APPARATUS.....	105
D. A POSA WOULD NOT HAVE BEEN MOTIVATED TO PERFORM CARRIER'S ALIGNMENT TECHNIQUE USING MANINIS.....	109
E. A POSA WOULD NOT HAVE BEEN MOTIVATED TO IDENTIFY INDIVIDUAL TEETH TO GUIDE IMAGE ACQUISITION IN EITHER CARRIER OR SALAH	124
F. A POSA WOULD NOT HAVE BEEN MOTIVATED TO CREATE A LEARNING BASE BY ALTERING THE METHOD DISCLOSED IN SALAH (CLAIMS 8 AND 9)	130
G. DR. FOROOSH'S SEMANTIC CONTOURS ARE NOT "IMAGE ATTRIBUTES" (CLAIM 7)	133
H. THE GROUND REFERENCES FAIL TO DISCLOSE THAT THE "ANALYSIS IMAGE IS ANALYZED AND CHECKED IN REAL TIME" (CLAIM 14)	140
VIII. CONCLUSION.....	141

EXHIBIT LIST

Exhibit Number	Description
2002	Declaration of John Mongan, dated December 7, 2023
2003	Curriculum Vitae of John T. Mongan, MD, PhD
2004	Redline of Carrier against the Carrier Provisional
2005	GeForce GTX TITAN X Specifications
2006	Andri et al., <i>YodaNN: An Architecture for Ultralow Power Binary-Weight CNN Acceleration</i>
2007	Shrestha and Dang, <i>Deep Learning-Based Real-Time Auto Classification of Smartphone Measured Bridge Vibration Data</i>
2008	<i>Dental Monitoring SAS v. Align Technology, Inc.</i> , Case No. 22-cv-07335, Dkt. No. 59, Order Denying Second Motion to Dismiss (N.D. Cal., July 19, 2023)
2009	<i>Dental Monitoring SAS v. Align Technology, Inc.</i> , Case No. 22-cv-07335, Defendant's Patent Local Rule 3-3 and 3-4 Initial Invalidity Contentions and Disclosures (N.D. Cal., September 11, 2023)
2010	<i>Dental Monitoring SAS v. Align Technology, Inc.</i> , Case No. 22-cv-07335, Dkt. No. 86, Align's Notice of Pending <i>Inter Partes</i> Review Petition (N.D. Cal., October 5, 2023)
2011	<i>Dental Monitoring SAS v. Align Technology, Inc.</i> , Case No. 22-cv-07335, Dkt. No. 92, Order Re Updated Patent Showdown Deadlines (N.D. Cal., October 26, 2023)
2012	<i>Dental Monitoring SAS v. Align Technology, Inc.</i> , Case No. 22-cv-07335, Dkt. No. 56, Patent Showdown Scheduling Order (N.D. Cal., July 14, 2023)
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.
2039	J. Liu et al., <i>Machine Learning in Orthodontics: Challenges and Perspectives</i> , 30(10) Adv. Clin. Exp. Med. 1065 (2021)

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
2043	M. Owen, <i>How iPhone Speeds Have Grown in the Last 5 Years</i> , Apple Insider (Sept. 26, 2022), https://appleinsider.com/articles/22/09/26/how-iphone-speeds-have-grown-in-the-last-5-years
2044	Apple, <i>iOS 11 is Available Tomorrow</i> , Apple Newsroom (Sept. 18, 2017)
2045	B. Lang, <i>NVIDIA's Next-Gen Titan X Has 11 TFLOPS Performance, 12GB of Memory, and Costs 1 Arm and 1 Leg, Road TOVR</i> (July 21, 2016)
2048	Deposition Transcript of Dr. Hassan Foroosh, Ph.D., in IPR2023-01269 (May 17, 2024)

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. 10,755,409 (“the ’409 patent”) (EX1001) and on matters relating to the patentability of the claims of that patent, and to respond to the opinions of Dr. Hassan Foroosh (EX1002). The following is my written declaration on these issues.

II. BACKGROUND

A. Education and Industry Experience

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

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

4. 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.

5. 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.

6. 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 4th 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. I have offered testimony from the perspective of a POSA regarding the '409 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

19. 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

20. 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 and those filed by Petitioner and Patent Owner in this proceeding.

21. 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.

22. 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.

III. UNDERSTANDING OF THE LAW

23. 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

24. 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 '409 patent in particular is summarized below.

B. Date of Invention

25. 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).

26. 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

27. 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

28. 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.

29. 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.

30. 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.

31. 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

32. 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 '409 patent issued was filed after March 16, 2013.

33. 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 '409 patent claims are unpatentable by a preponderance of the evidence.

F. Obviousness

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

G. Secondary Considerations Relevant to Nonobviousness

45. I understand that factors commonly known as “secondary considerations” may weigh against finding claims obvious. I understand that, for such secondary considerations to be accorded substantial weight, a nexus or causal connection must exist between the evidence and the merits of the claimed invention itself. I understand that, where the offered secondary consideration actually results from something other than what is both claimed and inventive in the claim, there is no nexus to the merits of the claimed invention. I further understand that such secondary considerations are not determinative, and that one may conclude that a patent claim is obvious even in the face of considerable evidence of secondary considerations. For example, such secondary considerations must be considered in light of all other factors and evidence that support obviousness.

46. I understand that secondary considerations may include:

- *Third Party Recognition:* I understand that evidence of praise of the invention by persons skilled in the art can be considered as an indication of nonobviousness where there is a nexus between the claimed invention and the praise. I understand that for there to be a nexus, the praise by persons skilled in the art must be due to what is both claimed and inventive in the claim. I understand that, at the very least, a nexus requires that any products that are praised fall within the subject matter of each claim.

- *Commercial Success*: I understand that evidence of commercial success of the invention can be considered as an indication of nonobviousness if there is a nexus between the claimed invention and the commercial success. I understand that for there to be a nexus, the commercial success must be due to what is both claimed and inventive in the claim. For commercial success based on sales, in particular, I understand that the sales must be a direct result of the unique characteristics of the claimed invention.
- *Licensing*: I understand that evidence of commercial success can also take the form of licenses for the claimed invention, but that such evidence likewise requires a nexus with the claimed invention. For example, I understand that the mere fact of licensing, without more, is generally not a strong indication of nonobviousness if it cannot also be shown that the licensees did so out of respect for the patent rather than to avoid litigation expense.
- *Failure of Others*: I understand that evidence that others tried but failed to solve the problem solved by the claimed invention can be considered as an indication of nonobviousness. However, like the other factors, this failure of others must have a nexus to what is both claimed and inventive in the claim.
- *Long-Felt But Unsolved Need*: I understand that evidence that a claimed invention solved longstanding problems or fulfilled a long-felt need in an industry can be considered as an indication of nonobviousness. But I

understand again that any such evidence must be considered in light of all other factors and evidence that support obviousness.

- *Unexpected and Superior Results*: I understand that where a claimed invention results in unexpected results that provide improved properties that are greater than would have been predicted, those unexpected results may indicate that the claimed invention yielded more than predictable results. I understand that when unexpected results are used as evidence of nonobviousness, the results must be shown to be unexpected compared with the closest prior art. I understand that any such evidence must be considered in light of all other factors and evidence relevant to obviousness.
- *Initial Skepticism and Rapid Replacement of Prior Art Technology*: I understand that evidence that a claimed invention faced initial skepticism from those of ordinary skill in the art, and that despite the initial skepticism, over time the claimed invention was widely adopted and praised, this evidence can be an indication of non-obviousness. As elsewhere, I understand that any such evidence must be considered in light of all other factors and evidence that support obviousness. As with other secondary considerations, it must be shown that the skepticism—and the eventual adoption and praise—result from what is both claimed and inventive in the claim.

- *Copying*: I understand that evidence of copying of the patented invention may be an indication of non-obviousness, but that any such evidence must be considered in light of all other factors and evidence that support obviousness. I further understand that not every competing product that arguably falls within the scope of a patent is evidence of copying; otherwise, every infringement suit would automatically confirm the nonobviousness of the patent. I understand that copying requires evidence of efforts to replicate a specific product, which may be demonstrated through internal company documents, direct evidence such as disassembling a patented prototype, photographing its features, and using the photograph as a blueprint to build a replica, or access to the patented product combined with substantial similarity to the patented product.

IV. THE '409 PATENT

A. The Effective Filing Date of the '409 Patent

47. The '409 patent was filed as U.S. Application No. 16/031,201 on July 10, 2018, claiming priority to French Patent Application No. 1756947, July 21, 2017. I understand that Petitioner has not challenged the entitlement of the '409 patent to this priority date. Pet. 13; EX1002, ¶26.

B. Person of Ordinary Skill in the Art

48. 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 multidisciplinary team, which would have included individuals familiar with medical imaging, dental imaging, dentistry, and/or orthodontics.

49. 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, ¶70.

50. 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*.¹

51. As Dr. Foroosh recognizes, Salah and Carrier (and the Carrier Provisional) are in the field of dental imaging and in several respects are directed to and/or assume the knowledge of dental professionals. EX1002, ¶81; EX2048, 34:10-13; *see also infra* §§IV., V.A-C., VI-VII. As such, it is my opinion that it would be "appropriate" for a POSA to consult with individuals having experience in dentistry and/or orthodontics to understand the teachings of Salah and Carrier (and the Carrier Provisional). Accordingly, I have consulted with Dr. Budi Kusnoto, D.D.S., M.S. regarding the dental and/or orthodontic aspects of the Salah and Carrier references, as well as the Carrier Provisional, in forming my opinions herein, including as specifically noted throughout my report.

C. Introduction to the '409 Patent

52. The '409 patent is directed to inventions relating to dental treatment. It addressed an "ongoing need for a method simplifying the analysis of" dentition, aligner, and other related images used to provide dental treatment to patients, *e.g.*, orthodontic treatment using aligners. *See, e.g.*, '409 patent, 1:16-18, 3:60-63, 4:21-

¹ All emphasis is added unless noted otherwise.

27. The patent solved that need by, for instance, providing new and improved methods that employ a deep learning device to acquire and analyze such images. *See, e.g., id.*, 1:22-28, 18:56-61, 28:44-48. The '409 patent discloses two separate image analysis methods: (a) a detailed analysis method, and (b) a global image analysis method. I have described each of these methods below.

53. In the “**detailed analysis method**,” step 1) is to create a learning base of historical images of dental arches that can be used to train the deep learning base. *Id.*, 15:34-16:12; *see also id.*, 1:29-40 (describing “Analysis by Tooth” as a “detailed analysis of an image called ‘analysis image’ of a dental arch of a patient” including the steps of “1) creation of a learning base comprising more than 1000 images of dental arches, or ‘historical images’” and “2) training of at least one deep learning device, preferably a neural network, by means of the learning base”).

54. “To create a historical image of the learning base” according to the detailed analysis method, for example, “an operator, preferably an orthodontist, identifies one or more ‘historical’ **tooth zones** on an image . . . then assigns[] to each . . . **a value for at least one tooth attribute.**” *Id.*, 15:34-16:12 (emphasis added); *see also id.*, 1:29-2:41 (describing “detailed analysis method” as one in which the learning base comprises historical images labeled with “one or more zones each representing a tooth, or ‘historical tooth zones’, to each of which, for at least one tooth attribute, a tooth attribute value is assigned”), 16:6-12 (disclosing that the

learning base comprising historical images labeled with tooth zones and tooth attribute values “may ... be constructed according to other methods”). The ’409 patent discloses that the “tooth attribute is preferably chosen from a tooth number, a tooth type, a tooth shape parameter, ... a tooth appearance 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.*, 8:9-19.

55. “In step 2), a deep learning device, preferably a neural network, is trained with the learning base.” *Id.*, 16:13-14; *see also id.*, 1:39-40. In this step, “the deep learning device is preferably trained by a learning process called ‘deep learning’. By presenting, as input for the deep learning device, historical images (images+descriptions), the deep learning device progressively learns to recognize patterns on an image, and to associate them with tooth zones and with tooth attribute values, for example tooth numbers.” *Id.*, 16:44-50.

56. Next, in step 3), “the image that is to be analyzed, or ‘analysis image,’ is submitted to the deep learning device,” which “[t]hrough its training in the step 2), ... is capable of analyzing the analysis image and of recognizing said patterns therein.” *Id.*, 16:51-55. In particular, the deep learning device may “determine a probability relating to: the presence, at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone’, the attribute value of

the tooth represented on said analysis tooth zone.” *Id.*, 16:55-60; *see also id.*, 1:41-46. For example, the deep learning device trained by a learning base according to the detailed analysis method is capable of determining that “there is a 95% chance that a form of the analysis image represents an incisor.” *Id.*, 16:62-65.

57. Next, in “step 4), the results of the preceding step are analyzed to determine the teeth represented on the analysis image.” *Id.*, 17:1-3; *see also id.*, 1:47-50 (“determination, as a function of said probability, of the presence of a tooth of said arch at a position represented by said analysis tooth zone, and of the attribute value of said tooth”). For example, if a probability associated with an analysis tooth zone and with a tooth attribute value for this analysis tooth zone determined by the deep learning device in step 3) exceeds a threshold, a determination can be made that “the analysis tooth zone effectively represents a tooth having said tooth attribute value.” *Id.*, 17:4-11. “The step 4) thus leads to the definition of an analysis image enriched with a description defining the analysis tooth zones and, for each analysis tooth zone, the values of the attributes of the tooth represented by the analysis tooth zone.” *Id.*, 17:12-15.

58. As the ’409 patent teaches: “The detailed analysis method described above advantageously allows for a fine analysis of the analysis image.” *Id.*, 3:19-21. The tooth attribute value also “may be used to define a value for an image

attribute,” which is a value that “relates to the whole image and not to part of the image.” *Id.*, 3:25-26, 58-59.

59. “Alternatively, the deep learning device may be used globally.” *Id.*, 3:22-23. The “**global image analysis method**” comprises steps a' to c' and is “similar to the detailed analysis method described above, apart from the fact that, according to the global analysis, it is not necessary to analyze the individual situation of each tooth.” *Id.*, 17:20-23; *see also id.*, 3:65-4:13. Rather, the “analysis is global to all the image” and “the deep learning device determines the value of an ‘image’ attribute without having to previously determine tooth attribute values.” *Id.*, 17:23-26. This method requires a different learning base to train the deep learning base. *Id.*, 3:22-25 (“Alternatively, the deep learning device may be used globally, the learning base containing historical images whose description provides a global attribute value for the image.”), 3:49-51 (“When the image attribute refers to the content of the image, the description of the historical images of the learning base specifies a characteristic of this content.”). Specifically, a learning base that comprises historical images of dental arches “include[ing] a description specifying an image attribute value for each image,” rather than a tooth zone labelled with a tooth attribute as in the detailed analysis method. *Id.*, 17:32-33; *see also id.*, 3:25-28 (“In other words, the value of the image attribute relates to the whole image and

not to part of the image. The attribute is not then a ‘tooth’ attribute, but is an ‘image’ attribute.”).

60. “Through its training” according to the global image analysis method, “the deep learning device is capable of analyzing the analysis image and of recognizing said patterns therein,” and “[b]ased on these patterns, it may in particular determine a probability relating to the value of the *image attribute* considered.” *Id.*, 17:37-41; *see also id.*, 4:7-12 (the analysis image step comprising “submission of the analysis image to the deep learning device for it to determine, for said analysis image, at least one probability relating to said image attribute value, and determination, as a function of said probability, of a value for said image attribute for the analysis image”). As the ’409 patent explains, “[t]he image attribute also makes it possible to detect, for example, whether the mouth is open or closed, or, more generally, whether the image is suitable for a subsequent treatment, for example whether it makes it possible to monitor the occlusion.” *Id.*, 3:31-36; *see also id.*, 4:13-19.

61. “The image attribute may in particular relate to a position and/or an orientation and/or a calibration of an acquisition apparatus used to acquire said analysis image, and/or a quality of the analysis image, and in particular relating to the brightness, to the contrast or to the sharpness of the analysis image, and/or the content of the analysis image, for example to the representation of the arches, of the

tongue, of the mouth, of the lips, of the jaws, of the gums, of one or more teeth or of a dental, preferably orthodontic, appliance.” *Id.*, 3:37-48.

62. As the ’409 patent teaches: “a global image analysis method according to the invention advantageously makes it possible to immediately and globally assess the content of the analysis image.” *Id.*, 4:32-34.

63. The ’409 patent further discloses that the detailed and global analysis methods provide a new and improved method of “embedded monitoring” for the acquisition of dentition, aligner, and other related images such that the process is more efficient and meets certain acquisition conditions—which ultimately is necessary for effective dental treatment. *See, e.g., id.*, 23:46-48 (“An image analysis according to the invention is also useful for guiding the acquisition of an image of a dental arch, in particular for establishing a remote diagnosis.”), 23:48-24:32, 25:13-20. For instance, by specifically determining whether the quality or another characteristic of an acquired image (*e.g.*, “opening of the mouth” or “wearing of a dental appliance”) meets a particular “setpoint” for such an image, requiring adjustments if that “setpoint” is not met, and “guid[ing]” an operator to make the correct adjustments, the ’409 patent invention can “considerably speed up the execution” of this method. *Id.*, 27:16-22, 28:44-48, 32:13-33. As another example, where the deep learning device has been trained to recognize analysis tooth zones and tooth attribute values, it can take advantage of that analysis to determine the

general orientation of a photograph (e.g., “left photo” or “right photo”) and determine whether that orientation complies with how the patient was instructed to acquire the image. *E.g., id.*, 26:9-25. The ’409 patent therefore provides a specific technological advance that provides for remote acquisition (including by a patient without medical training) and analysis of dentition, aligner, and other related images for the purposes of dental treatment.

64. As discussed above, in the embodiment according to the *detailed* analysis method, the learning base comprises historical images of dental arches “comprising one or more zones each representing a tooth, or ‘historical tooth zones’, to each of which, for said tooth attribute, *a tooth attribute value is assigned.*” *Id.*, 24:39-44. And in this embodiment, the method of embedded monitoring comprises “submission of the analysis image to the deep learning device so that it determines at least one probability relating to: the presence, at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone’, the attribute value of the tooth represented on said analysis tooth zone,” and “determination, as a function of said probability, of the presence of a tooth at a position represented by said analysis tooth zone, and of *the attribute value of said tooth.*” *Id.*, 24:47-58. In other words, the learning base is trained on images labeled with tooth attribute zones and tooth attribute values, and it is used in an analysis to determine tooth attribute 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.*, 24:61-25:4. “Advantageously, such a method makes it possible to directly obtain a value for an image attribute, without having to determine values for a tooth attribute” and “[i]t is therefore advantageously more rapid.” *Id.*, 26:27-31.

65. While the '409 patent contemplates the identification of analysis tooth zones (*e.g.*, tooth contours), a POSA would have recognized that a particular advantage of applying a deep learning device to the analysis image as described in the patent is the ability to identify complex and wholistic attributes of the teeth or the image, such as whether the patient's bite is open or closed or whether the “the dental situation ‘is pathological.’” *E.g., id.*, 3:19-36. An additional advantage of the embedded monitoring approach is that it allows for the acquisition of useful images by patients, who may otherwise make a “mistake in the acquisition of the updated images.” *Id.*, 25:34-52.

66. The '409 patent further discloses that updated images acquired by an operator, such as a patient, can be annotated with information from a 3D model of the patient's teeth (originally acquired during an in-office appointment and subsequently updated using the method described in the patent) and can form additional “historical images” to enrich the learning base on which the deep learning device is trained, reducing the expense of manually annotating such images. *E.g.*,

id., 2:4-3:17, 15:23-33 (an alternative is for an “operator” to “analyze[] thousands of analysis images”), Figs. 2-3, 5, 17. As the patent explains, for example, the “invention provides a method that is particularly effective for creating a learning base” that may be used to train the deep learning device to perform the analysis and make an accurate assessment, which is vital to effective dental treatment. *See, e.g., id.*, 13:62-63.

V. OVERVIEW OF THE ASSERTED ART

A. Salah

67. PCT Publication No. WO2016066651 to Salah et al. (“Salah”) is entitled “Method for Monitoring Dentition.” While Salah names five inventors, four of whom match the inventors listed on the ’409 patent, as discussed further below, it is my opinion that Salah is directed to an invention that is different from the ’409 patent.

68. 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. EX1008, 1:1-2:19; *see also* EX2002, ¶¶53-58.

69. 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 and “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.

70. 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 cell phone, 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

71. 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.

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

72. 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.

73. 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 cell phone is the only image acquisition apparatus disclosed in Salah that Dr. Foroosh relies on for his combination. EX1002, ¶¶47, 77, 92. 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.

74. In the following section, Salah describes the manner in which the patient is guided to acquire the “updated image.” In particular, Salah teaches that “the image-acquisition apparatus includes verification means to facilitate its

approximate positioning relative to the patient prior to acquisition of the updated image.” *Id.*, 27:19-20 (emphasis added). The need for only “approximate,” as compared to precise, positioning is consistent with Salah’s disclosure that the “updated image” is taken “without precise pre-positioning” (*id.*, 2:21-27) and which is a “considerable advantage” (*id.*, 31:23-24) of Salah’s invention.

75. Salah identifies several exemplary means of providing guidance to the patient such as via “written and/or voice prompts,” including ones that indicate whether the photo should be taken from the “front” or the “right,” by “displaying an arrow to orient the user,” or by indicating that the photo “is acceptable” or that the user “must take another photo.” *Id.*, 27:21-28:2. A POSA would have understood that such guidance is provided prior to (in the case of, *e.g.*, “take a photo from the front” or “take a photo from the right”) or during (in the case of, *e.g.*, the “signal to inform the user that the photo is acceptable or that, on the contrary, he/she must take another”) the photo image acquisition process, and prior to “completion of the acquisition process.” *Id.*, 27:21-25. Salah then goes on to give positioning as a specific type of guidance that can be provided during the image acquisition process by visual, audio, written, and/or vocal prompts. *Id.*, 27:25-28:2.

76. Salah subsequently describes the “verification means” for assisting with alignment of the “updated image.” *Id.*, 28:4-33:5. The “verification means” disclosed in Salah comprise a “reference,” which is a visual indicator that appears

on the viewfinder or screen of the patient's mobile device, such as a "line intended to be aligned with the general direction of the juncture between the upper and lower teeth" or marks representing where the patient's eyes, mouth, or face are to be positioned on the screen. *Id.*, 28:5-12. The "reference" is preferably in a fixed position on the screen. *Id.*, 28:11-12. Salah also indicates that the "reference" may be defined "at least in part" by a view of the reference model (*e.g.*, "augmented reality" in which the model of the patient's dental arch "in transparent form") which "makes it very easy *for the patient* to superimpose such a view approximately on the teeth to be photographed." *Id.*, 30:16-21 (emphasis added).

77. A POSA would have recognized that Salah teaches that the overlay of a model of the patient's dental arch (*e.g.*, a partially transparent image or outline of modeled teeth) is a "reference" designed to assist the *patient* with manual alignment, not as a guide for analysis with computer vision techniques. Salah never discloses that such an overlay could be used by the specialized application or computer program running on the acquisition apparatus, which Salah discloses may be used to automatically determine alignment and guide the patient accordingly. On the contrary, Salah expressly teaches away from using such comparisons to achieve alignment. *Id.*, 31:26-27 ("Preferably, no measurements are taken on the teeth to bring the acquisition apparatus into the matching position."). I understand from Dr. Kusnoto that, during orthodontic treatment, a patient's teeth may be moved several

millimeters. Thus, a POSA would have recognized that, in the context of the method for determining “the deformation and/or displacement of teeth between the steps a) and b)” of Salah, when the “updated image” is acquired, the patient’s individual teeth may be in an unknown position.

78. Salah next discloses the use of “registration marks” that correspond to and can be aligned with the “references” to get the desired alignment. For example, the “registration marks” can be attached to the patient, including preferably, on a dental retractor, and may take the form of shapes, colors, letters, or numbers, and may be of various sizes. *Id.*, 28:13-29:17. The use of “registration marks” on a retractor is illustrated in Figure 5a of Salah, where the “registration marks” are labeled “12.” *Id.*, 33:18-20. By matching the “registration marks” with the on-screen “references,” the user can get the desired alignment for acquiring the “updated image.” *Id.*, 29:18-23 (“The ‘match’ between a reference and a registration mark is a predefined arrangement of one in relation to the other. It indicates a particular positioning of the acquisition apparatus in relation to the registration mark. The match depends on the nature of the reference and the registration mark. The predefined situation, *which corresponds to target acquisition conditions*, may in particular be a total or partial superimposition, juxtaposition or alignment of the reference and the registration mark.” (emphasis added)).

79. Salah further discloses that its invention is “remarkable” in that it allows the target acquisition conditions to define the position of the “references” such that, when the “registration marks” on the retractor are aligned with the “reference,” an image can be acquired according to the target acquisition conditions.” *Id.*, 12:20-13:11. For example, “to measure the displacement of a tooth, the first target acquisition conditions may correspond to a positioning 40 cm from the retractor, in front of and at the height of the retractor” and/or “positioning at 40 cm from the retractor, at retractor height, at 45° to the right and left of the sagittal plane.” *Id.*, 13:20-23. A POSA would have recognized that, in the context of the method in Salah, in which the patient’s individual teeth are expected to move and may be in an unknown position relative to each other when the “updated image” is acquired, the patient’s teeth could not be reliably used as “registration marks” for aligning the acquisition apparatus so as to capture an “updated image” according to such target acquisition conditions. By contrast, the “registration marks” on a retractor would not be expected to move during the course of treatment.

80. Salah also discloses a “preferred embodiment” in which acquisition of the “updated image” is performed using an “acquisition kit,” which comprises a dental retractor with a “registration mark” and an image acquisition apparatus with “a screen for displaying an acquirable image,” *i.e.*, an image prior to acquisition, “and a computer program comprising program code instructions for displaying the

image on the screen, at least one reference on said screen[,] . . . said reference being preferably immobile on the screen, and arranged in a position known as the ‘match position’ in which, when ***the registration mark*** corresponds to the reference on the display, the scanned image shows the retractor at a predetermined viewing angle and/or distance.” *Id.*, 30:21-31:2 (emphasis added).

81. Salah later explains that “[p]referably, the acquisition apparatus is programmed so that in real time, to detect registration marks on the retractor, analyze their relative positions or dimensions and, as a result, inform the user of the device to modify the position of the acquisition apparatus in relation to the patient’s teeth.” *Id.*, 54:12-15; *see also id.*, 10:22-26 (“the first and/or second acquisition apparatus includes a computer program comprising program code instructions for locating the ***mark(s) on the retractor*** in real time, analyzing its/their position(s) and/or dimensions, in particular the relative position(s) of several mark(s) and, as a result, providing an indication, preferably light or sound, so as to inform the user of said acquisition apparatus” (emphasis added)). A POSA would have recognized that Salah ***only*** discloses that the application or computer program running on the acquisition apparatus itself would compare a “reference” to a “registration mark,” which does not include human facial features or tooth contours.

82. Salah never suggests that the features of the patient’s face or dental arch can be used as “registration marks” to be aligned with on-screen “references.”

Rather, as discussed above, a POSA would have understood that the “references” corresponding to the patient’s eyes, mouth, face, or dental arch are visually aligned by the patient during image acquisition, not analyzed by the application itself. *Supra* §V.A.2. A POSA would have recognized that it would be easier for a patient to align a larger and more familiar outline (*i.e.*, facial features) than it would be to align “references” with a set of “registration marks,” which may be smaller and may take more arbitrary forms that are helpful for subsequent image processing. *E.g.*, EX1008, Fig. 7. Salah also expressly states that “[p]referably, no measurements are taken on the teeth to bring the acquisition apparatus into the matching position.” *Id.*, 31:26-27. As discussed above, this is consistent with the disclosure in Salah that the “registration marks” may be used to capture updated images under target acquisition conditions by aligning those “registration marks” with an on screen “reference.” Because the patient’s individual teeth are expected to move and may be in an unknown position relative to each other when the “updated image” is acquired, a POSA would have recognized that they are not suitable for use as “registration marks” to ensure that the target acquisition conditions are satisfied.

83. Accordingly, a POSA would have understood that Salah teaches away from taking on the challenges—including the significantly more complex image processing tasks—associated with the evaluation of the imaged teeth or other facial features and that one should instead employ more easily detected and analyzed

“registration marks.” A POSA would have further recognized that, to the extent that the acquisition apparatus in Salah is providing feedback regarding whether the photo is “acceptable,” including by beeping in real-time “as the device’s positioning improves,” the specialized application or computer program running on that apparatus is analyzing the relationship between the “reference” and the “*registration marks*,” not any facial or dental features. *Id.*, 27:21-28:2.

84. As is made clear throughout Salah, the comparison of reference to registration mark is one that takes place in real-time. *See, e.g., id.*, 54:12-15 (“Preferably, the acquisition apparatus is programmed so that in real time, to detect registration marks on the retractor, analyze their relative positions or dimensions and, as a result, inform the user of the device to modify the position of the acquisition apparatus in relation to the patient’s teeth.”). Salah also repeatedly states that the software used to perform the real-time analysis of the image during image acquisition runs on the image acquisition apparatus itself. *Id.*, 10:22-26 (“the first and/or second ***acquisition apparatus includes a computer program*** comprising program code instructions for locating the mark(s) on the retractor ***in real time***, analyzing its/their position(s) and/or dimensions, in particular the relative position(s) of several mark(s) and, as a result, providing an indication, preferably light or sound, so as to inform the user of said acquisition apparatus” (emphasis added)), 16:13-21 (“a personal device, in particular a cell phone or tablet, in which such a program is

loaded”), 30:22-31:2, 54:12-15, and claims 4 and 7. Indeed, Salah states that “[i]n particular, the acquisition apparatus can be a cell phone, and the program can be a specialized cell phone app.” *Id.*, 32:1-12.

85. Importantly, Salah confirms that alignment during the image acquisition process need only be “approximate.” The method for “checking the positioning and/or shape of teeth according to the invention” in Salah uses “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, to accurately assess tooth displacement and/or deformation since the initial reference model was produced.” *Id.*, 2:21-27. Regarding the capture of dental images, Salah discloses: “The person creating the images also positions the image acquisition apparatus in an approximate manner.” *Id.*, 31:12-13, 31:26-27. “In particular, image acquisition can be carried out by the patient himself or by one of their relatives, using a simple cell phone.” *Id.*, 31:6-7. “The ability to acquire images with limited precision is a ***considerable advantage***, since it makes acquisition possible anywhere and by anyone.” *Id.*, 31:23-24 (emphasis added). Thus, Salah does not require or relate to the acquisition of highly accurate, high-quality dental images.

86. Finally, with respect to image acquisition, Salah teaches that the use of standardized “registration marks” in Salah has the additional advantage of indicating the “direction,” “distance,” and “orientation of the acquisition apparatus” after the

“updated image” has been acquired. *Id.*, 37:4-38:9. Also a registration mark “whose optical characteristics are known enables exposure time and sensitivity to be evaluated,” for example by comparing the brightness and/or color of the imaged registration mark with the characteristics of the registration mark under known acquisition conditions. *Id.*, 38:19-22. The “ratio between the area of the registration marks on the updated image and the area of the updated image is used to evaluate the focal length of the acquisition apparatus.” *Id.*, This is particularly important for the method in Salah, in which a subsequent step requires identification of the “acquisition conditions” under which the “updated image” was acquired, including the angle, positioning, and settings of the acquisition device. *Id.*, 32:26-33:5 (“[T]he retractor’s registration marks preferably have several functions... In step c), they also enable the updated images to be cropped. Lastly, the registration marks on the retractor, which appear on the images, in step d), allow the approximate virtual acquisition conditions approaching the actual acquisition conditions to be determined, which allows data processing to be accelerated.”).

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

87. 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. As noted above, 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.

88. 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.

89. Importantly, a POSA would have recognized that any generation of a contour map in Salah occurs after image acquisition is complete. It is not for the purpose of alignment during image acquisition, but instead, part of the process of updating the 3D model to track the position and/or shape of a patient’s teeth.

90. Salah identifies eighteen “preferred methods” and eighteen methods that are “not preferred” but “can also be used,” for determining “discriminant information.” 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. “Deep learning” is expressly identified at the bottom of the list of methods that are “not preferred.” EX1008, 36:14. 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, 199:11-200:4.

91. 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) and does not mention any other potential values. *Id.*, 18:14-15. Figure 6b is a simple contour map, 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.” In fact, none of the “preferred” methods would be appropriate for identifying and labeling the contours as associated with particular, individual teeth rather than simply identifying any and all contours apparent within the updated image as a whole (including, *e.g.*, the tongue, as appears to be visible in Fig. 6b).

92. Reference to the disfavored approach of “deep learning” would not have indicated to a POSA that Salah teaches using such a technique to identify which contours are associated with individual teeth. Particularly in view of Salah’s disclosure 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 mobile phone), a POSA—like the Salah inventors—would have been discouraged from using a resource intensive “deep learning” approach to determine a contour map. *See, e.g., id.*, 33:6-8, *id.*, 13:24-28 (“Preferably, steps (c), (d) and (e) are carried out by a person with no university training in orthodontics and/or outside any medical, dental or orthodontic practice, and/or without the use of a mechanical stabilization device for the acquisition apparatus and/or ***without the use of any device other than a cell phone*** . . . and/or without use of a standard calibration gauge.”) (emphasis added), *id.*, 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.”).

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

93. 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 two-dimensional 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.”). 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”—

the two images can be compared directly, and their differences should reflect how the patient's teeth have changed (*i.e.*, moved or deformed) since the model was last updated. *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.”).

94. 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.

95. 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.

96. 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.

97. Salah's discussion of contours as "discriminatory information" would not suggest to a POSA that contours of individual teeth would 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.*

98. 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.

5. Salah Does Not Disclose the '409 Patent Invention

99. In paragraph 43 of his declaration (EX1002), Dr. Foroosh states that "the '409 patent repeatedly states that Salah teaches many of the steps recited in its claims." Consistent with my opinions, however, the '409 patent never suggests that Salah discloses the methods claimed in the '409 patent in which the "analysis image" is analyzed "by means of a deep learning device trained by means of a learning base" so as to determine "a value for an image attribute," which is compared with "a

setpoint” and as a function of that comparison a specific information message is sent, as part of a method “to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image.” *Supra* §IV.C. As the ’409 patent explains, to “implement the method of [Salah], there may be a desire to acquire updated images from different acquisition directions, for example a front image, a right image and a left image.” EX1001, 25:22-24. “The operator, generally the patient, may however make a mistake in the acquisition of the updated images.” *Id.*, 25:34-35. “In particular, he or she may forget to take an updated image, for example the front view, or invert two updated images.” *Id.*, 25:35-36. “Typically, the operator may take an image on the right whereas a left image is expected from him or her.” *Id.*, 25:38-39.

100. “By virtue of the invention [in the ’409 patent, however], each updated image is an analysis image which may be analyzed and checked, preferably in real time.” *Id.*, 25:50-52. Specifically, steps a’) to c’) of the ’409 patent, which comprise activation of an image acquisition apparatus, analysis of the analysis image by means of a specifically trained deep learning device, and determination of a value for an image attribute, “make it possible to characterize the analysis image.” *Id.*, 26:34-35. This “characterization of the analysis image makes it possible to guide the operator if the analysis image does not correspond to the expected image.” *Id.*, 26:36-38. For example, the ’409 patent makes it “possible to determine that the

updated image has been ‘taken on the right’ and compare this image attribute value with the instruction which had been given to the operator to take the updated image on the left.” *Id.*, 25:53-57. If “the attribute value of the updated image (image taken on the right) does not correspond to the instruction (acquire an updated image on the left), the acquisition apparatus may immediately alert the operator in order for him or her to modify the acquisition direction.” *Id.*, 25:57-62.

101. As discussed above, Salah describes a computer program running on the acquisition apparatus that detects the alignment of registration marks with a reference on the screen. EX1008, 10:22-26 (“the first and/or second acquisition apparatus includes a computer program comprising program code instructions for locating the mark(s) on the retractor in real time, analyzing its/their position(s) and/or dimensions, in particular the relative position(s) of several mark(s) and, as a result, providing an indication, preferably light or sound, so as to inform the user of said acquisition apparatus”), 54:12-15 (“Preferably, the acquisition apparatus is programmed so that in real time, to detect registration marks on the retractor, analyze their relative positions or dimensions and, as a result, inform the user of the device to modify the position of the acquisition apparatus in relation to the patient’s teeth.”). Salah does not discuss using a deep learning device to confirm alignment.

102. Furthermore, while Salah states that the image acquisition device may determine whether the “updated image” is acceptable, a POSA would have

understood, in the context of the method disclosed in Salah, this means that the “registration marks” are arranged in a “match position.” *See id.*, 30:21-31:2. Contrary to Dr. Foroosh’s opinion (EX1002, ¶137), Salah does not disclose or suggest a method for determining whether the content of the image is acceptable, *e.g.*, whether specific teeth are visible, whether the tongue is retracted or the mouth is open or closed, or whether the patient is wearing their dental appliance. Instead, as noted above, Salah teaches away from a method that analyzes the patient’s teeth during image acquisition. EX1008, 31:26-27 (“Preferably, no measurements are taken on the teeth to bring the acquisition apparatus into the matching position.”). And post-image acquisition, Salah at most suggests that deep learning is a non-preferred method for discriminant information. I therefore disagree that Salah teaches using deep learning to determine “values indicating whether the required teeth are all present or not” as suggested by Dr. Foroosh (EX1002, ¶120).

B. Carrier

103. U.S. Patent Publication No. 2021/0068923 to Carrier, et al. (“Carrier,” EX1005) is entitled “Methods and Apparatuses for Dental Images.” Carrier was filed on November 18, 2020, and published on March 11, 2021. Carrier claims priority to (a) Application No. 16/827,594 (EX1010), filed March 23, 2020, (b) Application No. 15/803,718 (EX1011), filed November 3, 2017, and (c) Provisional Patent

Application No. 62/417,985 (“Carrier Provisional,” EX1006), filed November 4, 2016.

1. Carrier Is Limited by the Disclosure of the Carrier Provisional, Which Contemplates Use by Dental Professionals Not Patients

104. There are significant and relevant differences between Carrier and the Carrier Provisional, and it is my opinion that the Carrier Provisional is directed to a fundamentally different invention from that claimed in claim 1 of Carrier. EX2004 (highlighting the significant differences between the Carrier Provisional and Carrier).

105. For example, Carrier includes disclosure relating to a method for guiding a patient through a user interface on a mobile device to take a first and second plurality of photos of their own teeth for transmission to a dental professional for evaluation. EX1005, [0153], Claim 1. In contrast, the Carrier Provisional only discloses a method for dental professionals, as opposed to the patient themselves, to use a mobile phone to capture “high quality dental images.” *See, e.g.*, EX1006, [0006]; *see also id.*, [0057] (“The on-screen overlay with the outline of teeth can help doctors to take quality dental photos using the mobile phone.”).

106. I have reviewed Dr. Foroosh’s cited evidence that the “providing a mobile application on a mobile device associated with a patient” limitation in Element [1.1] of claim 1 of Carrier is supported by the Carrier Provisional, namely,

[0054]-[0055], [0073]-[0075], [0094]-[0095], Figs. 5A, 8. EX1002, ¶¶54-56, 61. It is my opinion that, there is nothing in these paragraphs or the Carrier Provisional as a whole, that suggest the reference to a “user” in the context of the mobile phone app refers to a *patient*. To the contrary, in [0095], the Carrier Provisional describes an app which presents a list of patients, and the user selects or adds a specific patient. A POSA would not have understood a *patient* to have access to personal data for multiple patients, only a doctor would use that type of app. Consistent with my opinion, [0095] of the Carrier Provisional goes on to state that “the app may guide the doctor or clinician in taking a predetermined series of images, such as those shown in FIG. 5A.”

107. I have also reviewed Dr. Foroosh’s cited evidence that the “guiding the patient” limitations in claim 1 of Carrier are supported by the Carrier Provisional, namely, [0036], [0061], [0073]-[0075], [0085], Figs. 4A-5A, 8, 10 of the Carrier Provisional. EX1002, ¶¶57-58, 61. It is my opinion that, when read as a whole, these disclosures do not describe guiding a *patient* to capture dental images in sufficient detail that a POSA could reasonably conclude that the Carrier Provisional inventors had possession of such an invention. Instead, these disclosures, and the Carrier Provisional viewed as a whole, concern guiding a dental professional how to capture dental images of a patient. *See, e.g.*, EX1006, [0004] (“Some doctors and dental technicians may instead attempt to capture dental photos using a mobile

phone”), [0056] (“The camera is typically held by the user (*e.g.*, dentist, orthodontist, technician) and images are taken from various positions. The user must manually adjust the patient position and/or camera position.”), [0056] (“Therefore, doctors can be guided to take dental photos by the on-screen overlay, saving time and effort.”), [0057] (“The on-screen overlay with the outline of teeth can help doctors to take quality dental photos using the mobile phone.”), [0081] (“For example, if the image is too dark, or blurry, or out of focus, a warn message is shown on the screen to warn the doctors to retake the photo.”), [0095] (“Thereafter, the app may guide the doctor or clinician in taking a predetermined series of images.”). Additionally, I understand from Dr. Kusnoto that for at least some of the standard images contemplated in the Carrier Provisional it would be essentially impossible for a patient to position a mobile phone to capture the image while maintaining a view of the screen to determine alignment of the overlay with their own teeth.

108. Finally, to the extent claim 1 of Carrier encompasses a “computer-implemented method” that uses machine learning, it is my opinion that such a method is not disclosed in the Carrier Provisional. While Carrier expressly mentions the use of “machine learning,” none of that disclosure was present in the Carrier Provisional. EX2004. The Carrier Provisional contains only a general reference to the use of “edge detection image processing” or “image recognition algorithms,” in the context of software or application software running on the image acquisition

apparatus. EX1006, [0063] (“Once the apparatus (*e.g.*, software/application software) determines that the teeth are in alignment with the overlay, as shown in FIG. 3B, the screen may display an indicator... For example, the method or apparatus may determine alignment when a threshold amount/percentage of the edges of the patient’s teeth in the image are within a predetermined minimum distance of the overlay.... The edges of the teeth may be determined by any appropriate technique (*e.g.*, edge detection image processing, etc.).”), [0080]-[0081] (“Any of these methods and apparatuses may further include reviewing the captured image on the mobile phone to confirm image quality and/or automatically accept/reject the image(s).”). Given the context, the reference to an “image recognition algorithm” would not have indicated to a POSA that the inventors of the Carrier Provisional contemplated the use of a “deep learning” algorithm.

109. The fact that the invention claimed in claim 1 of Carrier is not adequately supported by the Carrier Provisional is also made evident by the substantial additions to the specification of Carrier. *See, e.g.*, EX1005, [0041], [0052]-[0054], [0095], [0097], [0135], [0154]-[0156], [0179]-[0181]; *see also* EX2004. These disclosures added to Carrier are significant additions because they relate to a patient capturing their own dental images in order to allow for remote screening outside of the supervision of a dental professional, using their own mobile

device, and the use of machine learning. These disclosures are also notably absent from the dental professional driven approach in the Carrier Provisional.

2. Overview of Carrier

110. Dr. Foroosh agrees that, to the extent that Carrier is considered to be prior art, only the disclosure that was also present in the Carrier Provisional should be considered. EX2048, 127:16-128:11. Dr. Foroosh confirmed that his opinions were based only on that overlapping disclosure. *Id.* Accordingly, throughout the remainder of my declaration, I use the terms “Carrier” and “Carrier Provisional” interchangeably, and where I include citations only to the Carrier Provisional, my opinions apply equally to the corresponding disclosure as it appears in Carrier unless otherwise indicated.

111. The Carrier Provisional teaches that in “dental and/or orthodontic treatment, a set of 2D facial and dental photos is often required in addition to 3D impressions.” EX1006, [0004]; EX1005, [0004]. According to the Carrier Provisional, such photographs were typically captured using a camera, such as DSLR camera, but such cameras were “expensive” and “may be difficult to align” with the patient’s teeth. *Id.* Accordingly, “some doctors and dental technicians” had “attempt[ed] to capture dental photos using a mobile phone,” which requires holding the mobile phone “close to the teeth” in order to get a “high level of detail.” *Id.* But this can result in “blurry” and “optical[ly] distort[ed]” images as a result of a “wide

angle camera lens.” *Id.* If the phone is held “too far” from the teeth, however, it results in a different problem—the “dental photos cannot meet the orthodontic standards.” *Id.* I understand from Dr. Kusnoto that photographs that “meet the orthodontic standards” would be understood to be a reference to the standard set of orthodontic photographs that was well-established in the dental field in 2017 and is depicted in, for example, Figures 9 and 10 of the Carrier Provisional. A POSA would have understood that the Carrier Provisional aims to address the identified challenges with obtaining such standard orthodontic photographs.

112. The Carrier Provisional describes methods of “capturing high quality dental images” with a mobile phone, “instead of using expensive and bulky SLR cameras.” *See, e.g.*, EX1006, [0050]-[0054], [0095]-[0099]. Those methods use an on-screen “overlay comprising an outline of teeth in a predetermined view” that can be aligned with the patient’s teeth. EX1006, [0006]-[0007]; EX1005, [0006]-[0008]. The Carrier Provisional further describes the specific predetermined views to which the overlays correspond, referring to a “set of dental images according to the orthodontic standards.” EX1006, [0008], [0015]; EX1005, [0009], [0016]. For example, such standard views include “an anterior view, a buccal view[,] an upper jaw view, or a lower jaw view.” *Id.* Such standard views are depicted in the Figure 5A, and 9-11H. EX1006, [0020], [0025], [0032], [0035], Figs. 5A, 9-11H; EX1005, [0021], [0026], [0068], [0071], Figs. 5A, 9-11H. I understand from Dr. Kusnoto that

the goal with such photographs is to capture photographs of a patient's dental arch from a specific standard set of views.

113. The Carrier Provisional also discusses “using image recognition algorithms to ensure that all required teeth are visible.” EX1006, [0081]; EX1005, [0127]. I understand from Dr. Kusnoto that, in the context of standard orthodontic photographs, the “required teeth” are those teeth that are visible when taking a photograph of a specific view. This is consistent with the discussion in the Carrier Provisional that “[t]he overlay with an outline of teeth in a predetermine view can further provide information such as required visible teeth in the predetermined view. EX1006, [0060]; EX1005, [0105]. The Carrier Provisional further gives the example of Figure 2, which “illustrates a buccal view overlay which shows what teeth should be visible in the view.” *Id.* A POSA would thus have recognized that the Carrier Provisional teaches that its overlay method can be used to “ensure that all required teeth are visible.” *See* EX1006, [0081]; EX1005, [0128].

114. The Carrier Provisional further discloses that the dental photographs can be captured “when the overlay *approximately* matches with the patient's teeth.” EX1006, [0006]-[0008]; EX1005, [0007]-[0008]; *see also* EX1006, [0010] (the method can further comprise “automatically triggering an indicator when the overlay approximates matches with the patient's teeth”); EX1005, [0011] (same); EX1006, [0060] (“an *approximate* visual match is needed to select the correct distance,

correct angle and correct view of the patient’s teeth”); EX1005, [0105] (same); EX1006, [0063] (alignment when only 50% to 90% of outline is within 0.1 mm to 1 mm of overlay); EX1005, [0108]; EX1006, [0011], [0019], [0022], [0034], [0050], [0056], [0059], [0061], [0065]-[0067], [0072], [0084]-[0085], [0089], [0105]; EX1005, [0012], [0020], [0023], [0070], [0095], [0101], [0104], [0106], [0110]-[0112], [0118], [0132]-[0133], [0166], [0186].

115. A POSA would have understood the Carrier Provisional’s references to “approximate” matches between the overlay and the patient’s teeth to be consistent with the goals of the Carrier Provisional, namely providing a method to capture a standard set of orthodontic photographs using a mobile phone consistently with the capabilities of “traditional dental photography” using a camera having a superior lens—images that I understand were often taken freehand without additional alignment tools. EX1006, Figs. 1A-C; EX1005, Figs. 1A-C.

116. The need for only approximate alignment to achieve the goals of the Carrier Provisional is consistent with its disclosure that a non-patient specific overlay can be used. EX1006, [0009] (“the overlay can comprise a *generic* overlay”); EX1005, [0010]. The Carrier Provisional indicates that the overlays can also be “modified according to the predicted or actual changes of teeth of the patient with treatment . . . thus matching the teeth of the patient more precisely.” EX1006, [0064]; EX1005, [0109]. I understand from Dr. Kusnoto that there can be significant

variability in patients' teeth, which could prevent even an approximate match with a generic overlay and thus explains why a "patient-specific" overlay might be appropriate. Kusnoto §§VI.A, VII.A. A POSA would not, however, have understood the ability to "match[] the teeth . . . more precisely" as displacing Carrier's disclosure that the resulting alignment need only be "approximate." *See* EX1006, [0033]-[0034], Fig. 3B (showing "patient-specific overlay" only "approximately matches with the patient's teeth," particularly for left- and right-most teeth); EX1005, [0069]-[0070], Fig. 3B (same).

117. The Carrier Provisional further discloses that the method can further comprise "automatically triggering an indicator when the overlay approximately matches with the patient's teeth." EX1006, [0010]; EX1005, [0011]. This method can further comprise "estimating an indicator of the distance between an edge of the patient's teeth at two or more regions and the outline of teeth and comparing that indicator to a threshold value," where the indicator "can be a visual indicator, such as a change of color," or "a voice indicator." *Id.*

118. Like Salah, the method disclosed in the Carrier Provisional concerns real-time guidance during actual image acquisition. *See, e.g.,* EX1006, [0054] ("In particular, these apparatuses and methods may include the use of an 'overlay' on a real-time image of the screen, providing immediate feedback on each of the desired orientations."); EX1005, [0007] ("For example, the methods described herein may

include displaying, on a screen (*e.g.*, of a mobile communications device such as a smartphone, etc.), an overlay comprising an outline of teeth in a predetermined view. The overlay may be displayed atop the view from the camera, which may show[] a real-time display of from the camera.”); [0099] (“In particular, these apparatuses and methods may include the use of an ‘overlay’ on a real-time image of the screen, providing immediate feedback on each of the desired orientations, which may also be used to adjust the lighting and/or focus, as described herein.”).

119. Also, like Salah, the Carrier Provisional discloses the use of software running on a mobile phone. EX1006, [0019] (“For example, described herein are non-transitory, computer-readable storage media storing a set of instructions ***capable of being executed by a processor of a mobile phone***, that, when executed by the processor, cause the processor to display real-time images of the patient’s teeth on a screen of the mobile phone, display an overlay comprising an outline of teeth in a predetermined view atop the images of the patient’s teeth, trigger an indicator when the overlay approximately matches with the patient’s teeth, and enable capturing an image of the patient’s teeth when then indicator is triggered.”); EX1005, [0020]; EX1006, [0063] (“Once the apparatus (*e.g.*, software/application software) determines that the teeth are in alignment with the overlay, as shown in FIG. 3B, the screen may display an indicator, such as, in FIG. 3B, changing the color of the overlay from the first color to a second color (*e.g.*, green, blue, etc.) and/or

automatically taking the image.”); EX1005, [0108]; EX1006, [0088] (“Various embodiments of the disclosure further disclose a non-transitory, computer-readable storage medium storing a set of instructions capable of being executed by a processor of a mobile telecommunications device...”); EX1005, [0165]; EX1006, [0089] (“For example, the non-transitory, computer-readable storage medium, wherein the set of instructions, when executed by the processor, can further cause the processor to estimate an indicator of the distance between an edge of the patient’s teeth . . . in the view of the patient’s teeth.”); EX1005, [0166]. While the images can be transmitted to a remote server and evaluated “for medical treatment,” there is no suggestion in the Carrier Provisional that any analysis on such a “remote server” is for the purposes of guiding the image capturing process. EX1006, [0014]; EX1005, [0155]-[0156].

120. While Carrier expressly mentions the use of “machine learning,” all such disclosure concerning “machine learning” was added after the priority date of the ’409 patent. EX2004. The Carrier Provisional contains only a general reference to the use of “edge detection image processing” or “image recognition algorithms,” in the context of software or application software running on the image acquisition apparatus. EX1006, [0063] (“Once the apparatus (*e.g.*, software/application software) determines that the teeth are in alignment with the overlay, as shown in FIG. 3B, the screen may display an indicator. . . . For example, the method or apparatus may determine alignment when a threshold amount/percentage of the

edges of the patient's teeth in the image are within a predetermined minimum distance of the overlay. . . . The edges of the teeth may be determined by any appropriate technique (*e.g.*, edge detection image processing, etc.)."); EX1005, [0108]; EX1006, [0080]-[0081] ("Any of these methods and apparatuses may further include reviewing the captured image on the mobile telecommunications device to confirm image quality and/or automatically accept/reject the image(s)."); EX1005, [0126]. The reference to an "image recognition algorithm" would not have indicated to a POSA that the Carrier Provisional contemplates the use of a "deep learning" algorithm, including the neural network taught in Maninis. Maninis discloses a much more complex architecture than is necessary to perform the method disclosed by the Carrier Provisional, and as discussed below, the Maninis neural network architecture is not suited to operate on the mobile devices used in the Carrier Provisional. That a POSA would not have understood the Carrier Provisional to refer to deep learning is confirmed by Carrier's addition of "machine learning" as an image recognition technique only in a subsequent Carrier patent application filed after July 2017, including by reference to "the machine learning framework provided by Apple with iOS 11," which first launched on September 19, 2017—after the

priority date of the '409 patent. EX1005, [0033], [0106], [0147]; EX2044 [Apple press release].²

121. While Dr. Foroosh confirmed in his deposition that the methods disclosed in the Carrier Provisional do not require determining individual tooth edges, Dr. Foroosh nonetheless suggests that Carrier discloses (1) determining the boundary for each individual tooth and (2) evaluating the distance between each such tooth-specific boundary and the corresponding tooth-specific region of an overlay. EX1002, [0086] (“the **tooth** edges Carrier uses to assess alignments”), [0109] (“Salah and Carrier teach performing analysis . . . to determine values for tooth presence and **tooth** edges”); [0112] (“the set of **tooth distances**”), [0114] (“the distance between teeth and overlay is calculated as a function of the edges (that is, the edges are used as an input to calculate the distances between **each tooth** and the overlay”), [0118] (“individual teeth corresponding to the contours used by Carrier to assess alignment”), [0121] (“a set of all tooth distances or an average of those distances”); EX2048, 113:21-116:2. I disagree.

² To be clear, the availability of machine learning capabilities on Apple iOS 11 would not impact my opinions regarding the Maninis neural network disclosures, which are far more resource intensive than the specialized, lightweight capabilities that Apple developed for use on its mobile devices.

122. A POSA would have recognized that the overlay in the Carrier Provisional is a binary bitmap representing a single contour comprising an outline of teeth matching the view in the desired orthodontic photo. Nothing in the Carrier Provisional suggests that the overlay comprises several bitmaps or semantic contours such that each part of the contour is recognized by the application as corresponding to a distinct tooth. Nor does the Carrier Provisional anywhere suggest that the determination of contours involves determining which portions of that contour corresponds to which individual tooth. In fact, a POSA would have understood that the comparison of the contour of the patient's teeth to the overlay does not require awareness that the detected contours even correspond to teeth at all, let alone which portions of that contour correspond to which teeth.

123. Accordingly, the Carrier Provisional never discusses measuring the distance between the part of a contour associated with an individual tooth and its corresponding portion of the contour associated with the targeted view. Carrier exclusively discusses the alignment of "teeth" or the "edges of teeth" *as a whole* as being compared to the overlay to determine sufficient alignment. *E.g.*, EX1006, [0063] ("the method or apparatus may determine alignment when a threshold amount/percentage of the *edges of the patient's teeth* in the image are within a predetermined minimum distance"); EX1005, [0108]. Dr. Foroosh's suggestion that the contours of individual teeth *could* be determined appears to be hindsight driven

by the '409 patent's requirement of the detection of "analysis tooth zones." Nor would a POSA read such a technique into Carrier's disclosure because a POSA would not have assumed that Carrier is implicitly suggesting the use of such an unnecessary, and resource-intensive evaluation.

C. Maninis

124. 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 below a high-level explanation of neural networks as they would have been understood by a POSA in July 2017.

1. Neural Networks

125. 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,

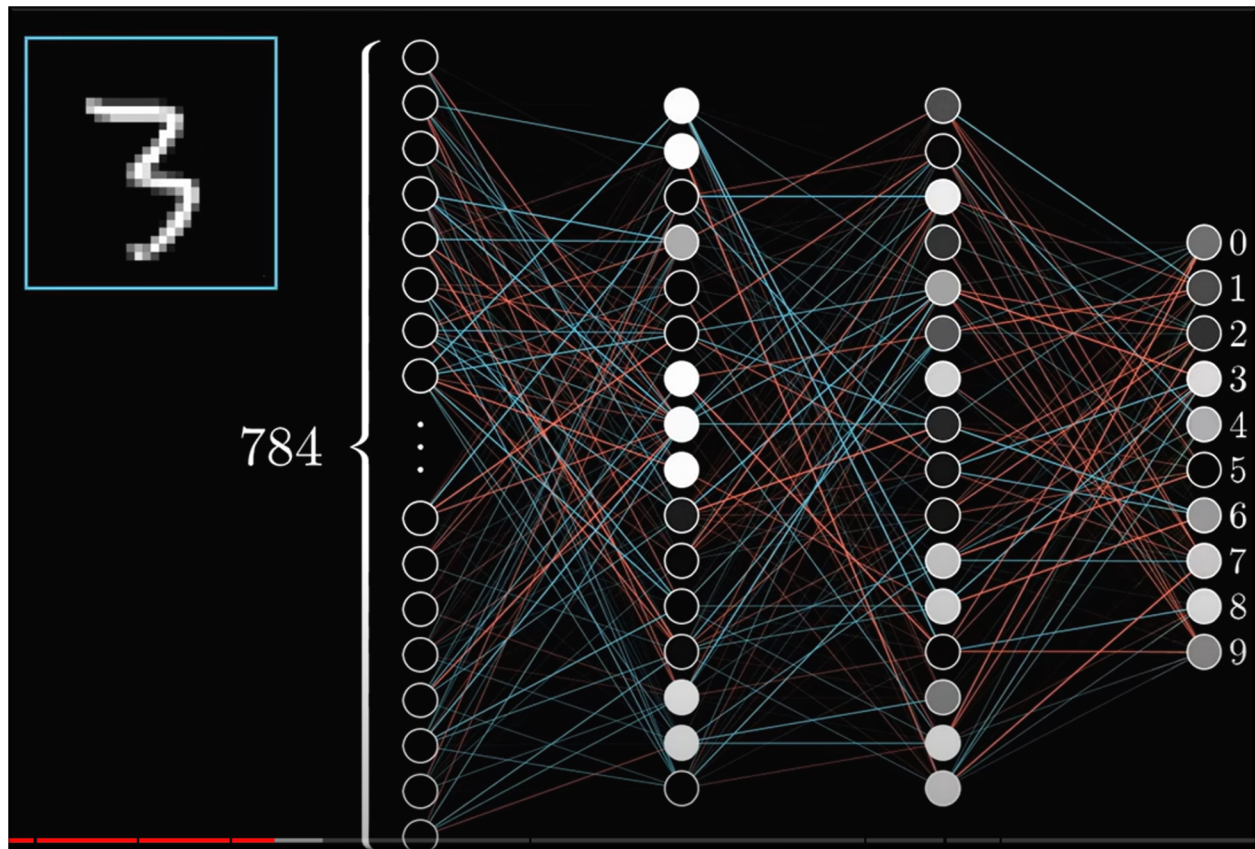
and can only recognize what they have been trained to recognize. In other words, it was known that each neural network trained for an image recognition task is unique and specific to the annotated images on which it has been trained, and its success (then and now) depends on the quality of that training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

126. 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> at 0:16-0:21(last accessed May 27, 2024).

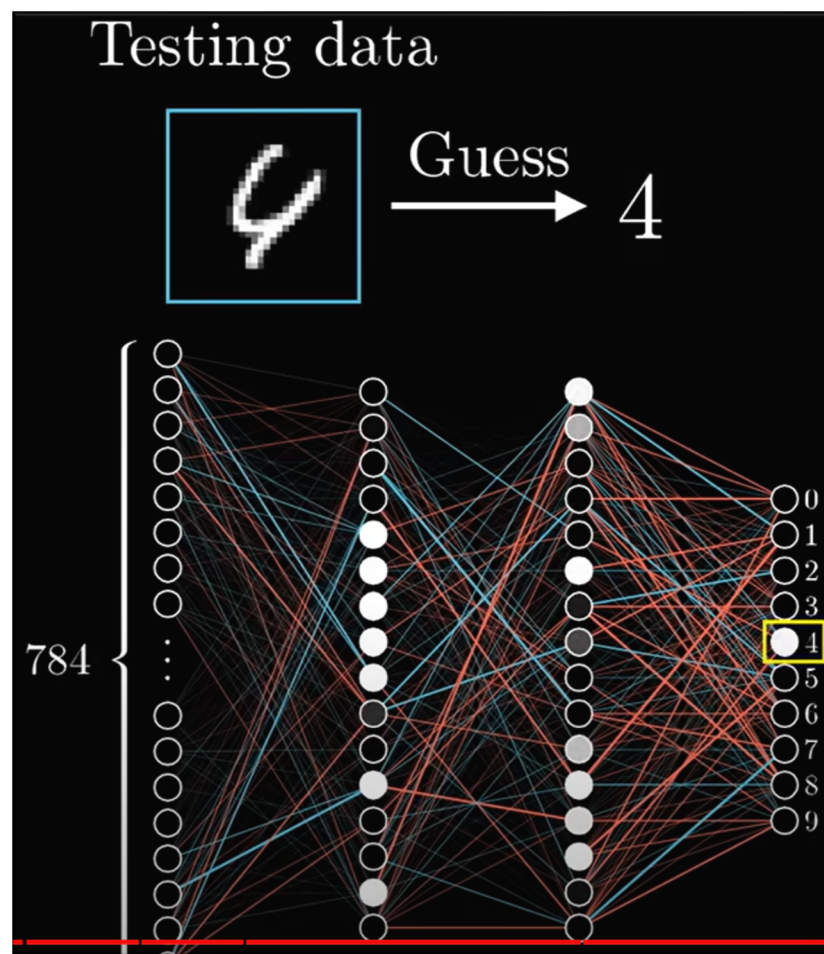
127. At the start of training, the neural network comprises a set of calculations using parameters called “weights” and “biases” that are initially set to random values and applied in several rounds of calculation or “layers,” with each influencing the next. The neural network might, for example, take 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=PLZHQObO WTQDNU6R1_67000Dx_ZCJB-3pi&index=2 at 3:34 (last accessed May 27, 2024).

128. 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.”

129. 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 at 2:34 (last accessed May 27, 2024).

130. While this is a simple example, the Maninis COB neural network is much more complex and has more layers.

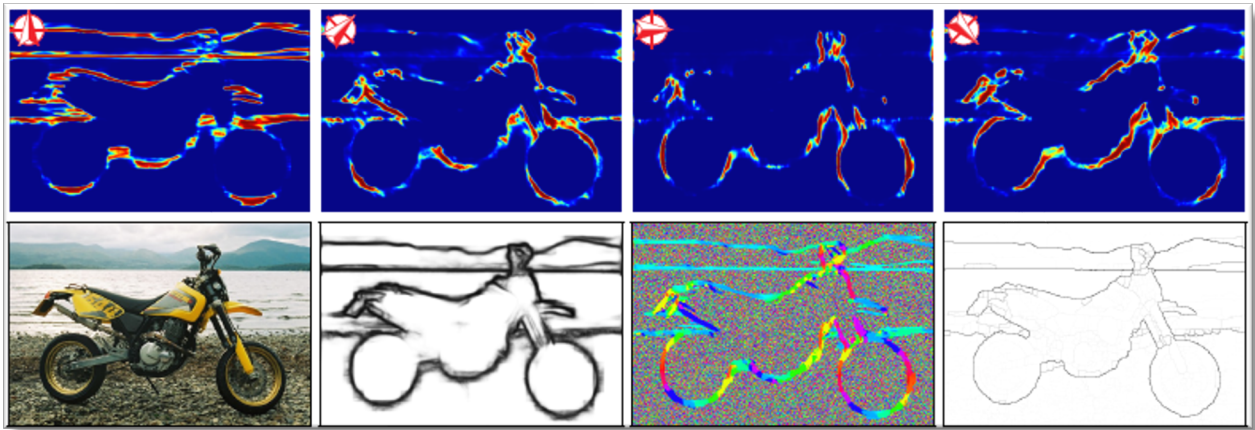
2. Overview of Maninis

131. Maninis is an academic paper and makes no mention of the use of its neural network architecture for analysis of dental or orthodontic images. Maninis also does not disclose that its neural network architecture could be used during image acquisition. 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.

132. 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.³ EX1007, 1.

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

133. 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) or 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. at 4, Fig. 3 (hierarchical regions depicted in bottom right image).

(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.

134. 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.

135. 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 (emphasis added). A POSA would have understood that "without distinguishing the semantics" means without identifying what object is contained within the boundary. This is illustrated by Figures 9 and 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).

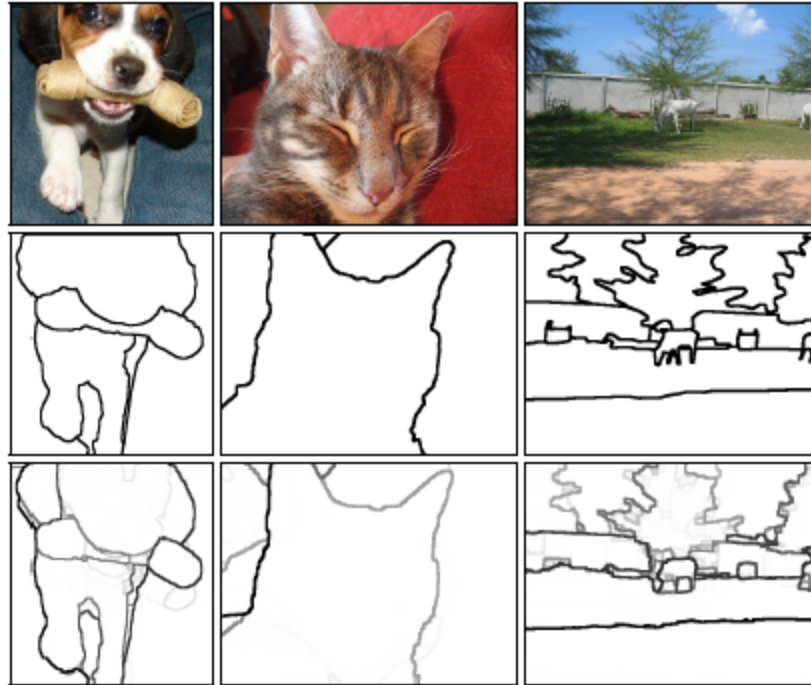
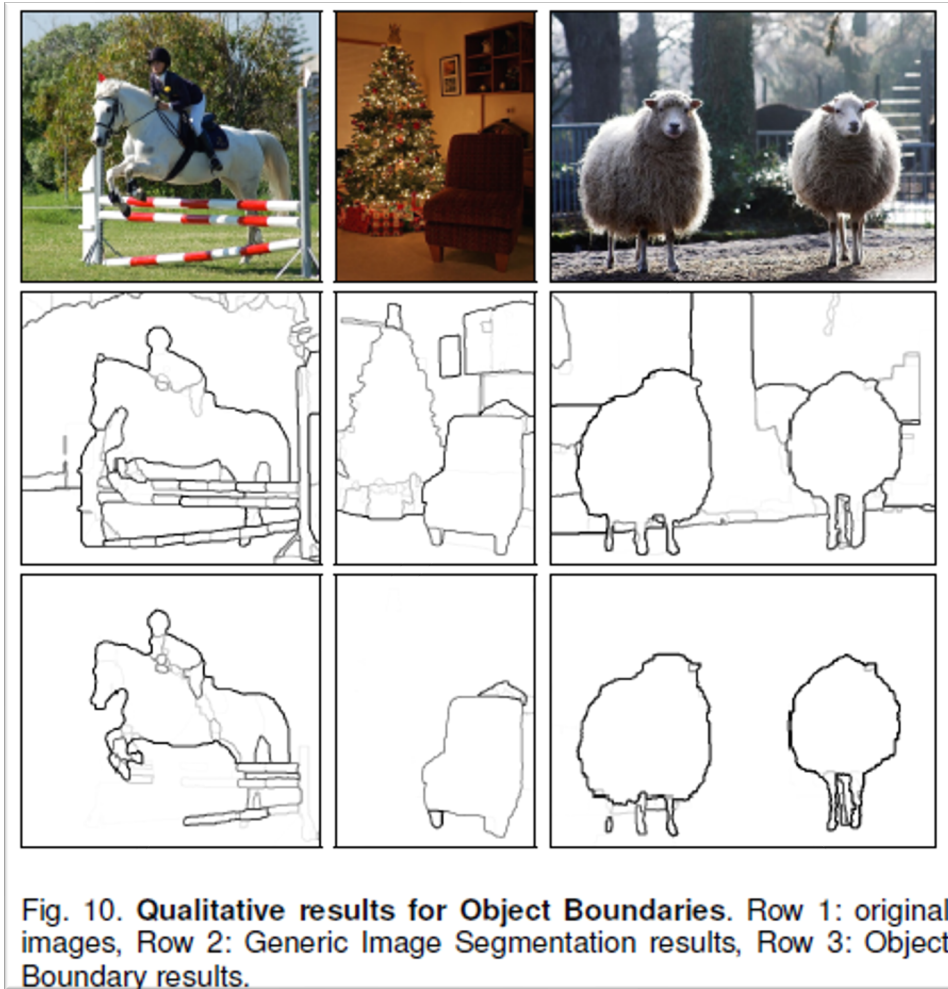


Fig. 9. Qualitative results on PASCAL - Hierarchical Regions. Row 1: original images, Row 2: ground-truth boundaries, Row 3: hierarchical regions with COB.

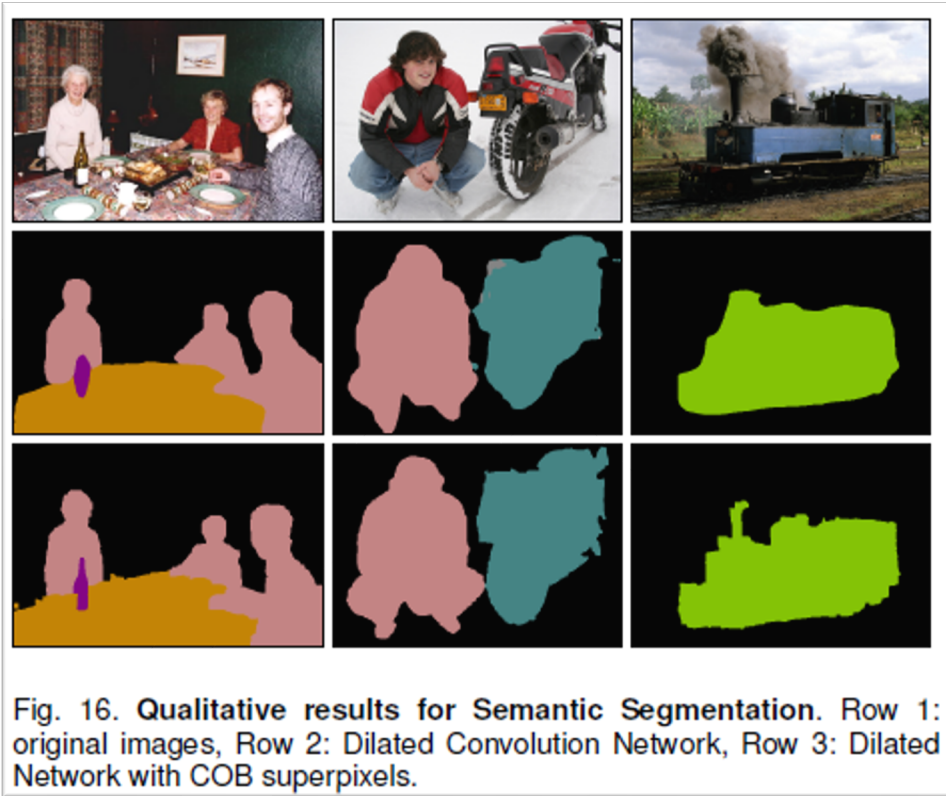
Although Maninis describes these contours as “semantically meaningful,” a POSA would have recognized that no “semantic segmentation” or object classification was performed by the Maninis COB neural network. EX1007, 7. Like Figure 9, Figure 10 illustrates “generic segmentation and object boundary detection,” 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 that these contours correspond to the specific objects they bound (*e.g.*, a horse, a chair, or a sheep). *See id.* This contrasts with Figure 16 (reproduced below), 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).

136. 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 “**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.

137. 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.

138. 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.

139. 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.

140. 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. But 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.

141. 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).

142. 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 graphics cards 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.

143. As of July 2017, a POSA would have known that the Maninis COB neural network was too computationally expensive and resource draining to run on even the most optimized mobile devices, and it was not suitable for real-time object recognition on devices with power limitations. *See, e.g.*, EX2006 [Andri], 48 (“Convolutional neural networks (CNNs) have revolutionized the world of computer vision over the last few years, pushing image classification beyond human accuracy. The computational effort of today’s CNNs requires power-hungry parallel processors or GP-GPUs. Recent developments in CNN accelerators for system-on-

chip integration have reduced energy consumption significantly. Unfortunately, even these highly optimized devices are above the power envelope imposed by mobile and deeply embedded applications”); EX2007 [Shrestha], 2 (confirming, after the ’409 patent priority date, that “[s]uch costly computation complexity makes the deployment of CNN models unaffordable for common PCs and mobile devices”).

144. The Nvidia TITAN X deployed for testing the Maninis COB neural network was an expensive, cutting-edge GPU in 2017 that is designed for use in desktop computers and was not compatible with mobile devices. EX2005 [GeForce] (identifying TITAN X GPU as having dimensions of 4.376 x 10.5 inches and a recommended power supply of 250 watts). In 2017, even the most advanced mobile phones—for example, the iPhone 8—had a processing power over an order of magnitude lower (neural engine processing capacity of 0.6 TFLOPS with 2 GB of memory) than the TITAN X GPU. *See* EX2043 [Apple Insider] (identifying FLOPS for iPhone 8, launched in September 2017). Dr. Foroosh has not explained how the Maninis COB neural network could be adapted to run on a mobile phone. A POSA in 2017 would have recognized that, even if Maninis’s COB neural network could be adapted to function on a mobile device available in 2017 (accounting for software compatibility), a conservative, back of the envelope calculation indicates that it would take more than 5 seconds to calculate a generic image segmentation

(assuming 100% utilization of the “neural engine” processing power on a leading mobile phone in July 2017) $(11/0.6)*0.28 = 5.1$). Although the Maninis authors fail to disclose the CPU used to compute the hierarchical segmentation (“OWT/UCM”), a POSA would have expected that the CPU in a desktop computer with a state-of-the-art graphics card would conservatively operate twice as quickly as the CPU of a state-of-the-art mobile phone. EX1007, Table 3 (noting with asterisk that only initial contours are run using GPU). It would thus conservatively take longer than 6 seconds to calculate hierarchical segmentations in one dental image on the iPhone 8 $(5.1+0.51*2=1.02)$. It would take longer—likely, significantly longer—to apply a second neural network to determine the semantic contours on which Dr. Foroosh’s combination depends, and to analyze multiple dental images.

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

145. 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, acquired with mobile phones, labeled with semantic contours of individual teeth by tooth number. EX1002, ¶167; 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).

146. A significant driver for the rising popularity of the application of deep learning devices to problems in other fields (outside of dentistry and orthodontics⁴) 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 dental images, acquired with mobile phones, labeled with semantic contours of individual teeth by tooth number that he relied on for his combination as existing in the prior art, and I am aware of none. Nor does Dr. Foroosh cite a single example of a neural network trained on such dental 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 dental 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. EX2039 [Liu], 1072 (emphasizing in 2021 that

⁴ Dr. Foroosh did not identify any prior art example of the application of deep learning in the dental field, and specifically, to the acquisition and analysis of dental photographs, and I am not aware of one.

“[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.”).

147. I understand from Dr. Kusnoto that it was not a regular practice for dental practitioners in 2017 to acquire dental images 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, in 2017, it would likely have taken over a year for even a large dental practice to acquire more than 10,000 dental images and thereafter, there would be a significant expense associated with de-identifying patient data for purposes of privacy laws such as HIPAA before they could be used in a learning base.

148. The annotation of images for training a neural network can also be a time-consuming and expensive process. I further understand from Dr. Kusnoto that the identification of individual tooth contours by tooth number in images would, in 2017, likely have required labeling by a dental professional trained to recognize teeth accurately. *See* EX2048 144:20-145:8 (indicating he could not identify the molars in a dental image, and indicating that a POSA would not have this knowledge). This

is consistent with my personal experience with the labeling of semantic contours in medical images in my practice as a radiologist and in connection with my own research, and that the specificity of these contours often requires trained professionals and can take a significant amount of time to label.

149. Accordingly, a POSA seeking a learning base of dental images labeled with semantic contours representing individual teeth, as proposed by Dr. Foroosh, would have recognized that the cost associated with such a learning base could range from tens of thousands to hundreds of thousands of dollars depending on total number of training images required to achieve sufficient results, and time associated with acquisition and labelling could also be significant.

VI. CLAIM CONSTRUCTION

A. “deep learning device”

150. I understand that Dr. Foroosh 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, ¶99. I also understand that Dr. Foroosh applied this construction in his declaration and has not formed his own opinion regarding the meaning of this claim term. EX2048, 71:6-73:19.

151. In my opinion, a POSA would have understood that, in the context of the '409 patent, the term “deep learning device” refers to a device that, through training, is capable of analyzing an image and recognizing patterns therein. EX1001, 16:53-55 (“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.*, 17:37-39 (same), 24:38-59. While I agree that a neural network is an example of a deep learning device, I have been asked to address the Board’s statement in the Institution Decision that “[t]he record before us is devoid of any evidence to substantiate Patent Owner’s inference that the ‘deep learning device’ recited in the claim is somehow integrated in the mobile device.” Paper 10 at 32. To the extent that the Board understood that the “deep learning *device*” refers to a physical device (*e.g.*, processor or computer), I respectfully disagree that a POSA would have understood the phrase “deep learning device” to have that meaning in the '409 patent.

152. The '409 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, 31:61-67. It further repeatedly states that the “deep learning device” is “preferably a neural network.” *E.g.*, EX1001, 1:22-28, 1:39-40, 1:66-67, 3:14-18, 8:41-43;

13:64-65, 16:14-43. The '409 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.*, 16:15-43. The '409 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.

153. 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.*, 31:61-67 (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 (confirming that a neural network can be described as “an algorithm with values that are updated through the training process” and clarifying that the “layers” in its “architecture” are “not ... hardware”), 209:24-210:5 (confirming that, “once ... trained,” neural networks can be “put ... on any number of computers and run[]”).

B. “real time”

154. Claim 14 recites the limitation “said analysis image is analyzed and checked in real time.” A POSA would have understood that the term “real time” means that the analysis and checking occur “immediately,” while the image acquisition process is still underway.

155. The '409 patent uses the term “real time” in the specification in the same context as the claim. Specifically, the specification gives an example of an image being “analyzed and checked ... in real time” where an updated image is analyzed and compared to the acquisition instructions (“updated image has been ‘taken on the right’”) sufficiently quickly to “immediately alert the operator . . . to modify the acquisition direction” if the “value of the updated image . . . does not correspond to the instruction.” EX1001, 25:50-62.

C. “tooth shape parameter”

156. The '409 patent indicates that the term “tooth shape parameter” is one type of “tooth attribute,” along with other examples such as “tooth number,” “tooth type,” “tooth appearance parameter,” “a parameter relating to the state of the tooth,”

or “an age for the patient.” EX1001, 8:9-21, cl. 4. Examples of “tooth shape parameters” in the ’409 patent are “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.” *Id.* Importantly, these “tooth attributes” are associated with but distinct from “tooth zones,” the latter of which is a “zone representing, at least partially, [a] tooth in [an] image.” *E.g., id.*, 1:30-50, 6:30-35, 13:31-46, 16:6-12, 16:44-50.

157. A POSA would accordingly have understood that the “tooth zone” (*e.g.*, “analysis tooth zone”) is a zone in an image corresponding (at least partially) to a tooth (or the visible part of a tooth) depicted therein. The “zone” would be understood as defined by a boundary or contour defining the edges of the tooth depicted in the zone. The “tooth attribute,” including a “tooth shape parameter,” describes the tooth depicted within that boundary (*e.g.*, its width).

158. I disagree with Dr. Foroosh’s apparent interpretation of the term “tooth shape parameter” as encompassing the boundary or contour defining the analysis tooth zone. *E.g.*, EX1002 ¶164. A contour is not descriptive—it is a map of pixels outlining the border between an object depicted in the image and other objects/background. It is itself equivalent to a zone in an image that can be described. A “tooth shape parameter” is a parameter of the shape of the tooth that is depicted *within that contour*, consistent with the ’409 patent examples of the tooth’s

width, thickness, height, etc. Dr. Foroosh's interpretation improperly conflates "tooth zone" and "tooth attribute." Accordingly, I disagree with each instance in which Dr. Foroosh relies on a tooth contour as disclosing a "tooth shape parameter" or "tooth attribute value."

VII. THE '409 PATENT WOULD NOT HAVE BEEN OBVIOUS IN VIEW OF SALAH, CARRIER, AND MANINIS

159. For the reasons set forth below, I disagree with Dr. Foroosh that the challenged claims of the '409 patent would have been obvious in view of Salah, Carrier, and Maninis.

- 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 be trained and would not perform in the proposed manner. Dr. Foroosh's proposed combination is thus inoperable and does not disclose the limitations of the '409 patent claims nor would a POSA have been motivated to make such a combination with a reasonable expectation of success.
- A POSA would not have been motivated to combine Carrier's alignment technique with Salah's methods and apparatus. Salah already discloses an alignment technique, and a POSA would have understood that displacing it would be detrimental to Salah's operation.

- A POSA would not have been motivated to use Maninis to perform the Carrier' alignment technique. Even if operable, Maninis is unfit for the purpose of Carrier's technique, which requires real time analysis on a mobile device. Additionally, a POSA would be discouraged from using Maninis because of the significant complexity and expense that would be incurred for no (or minimal) benefit.
- A POSA would not identify individual teeth to align the acquisition apparatus as Dr. Foroosh proposes. Counting teeth does not assist with the alignment of dental images, and Salah expressly discourages deploying such a technique.
- A POSA would not have been motivated to use 3D models to create a learning base as required in Claims 8 and 9. Dr. Foroosh cites no precedent for this approach, and a POSA would have understood that such an approach would be complex and expensive.
- Dr. Foroosh's proposed combination does not disclose the limitations of claim 7. Dr. Foroosh proposes that the learning base comprise historical images labelled with semantic contours, which are not "image attributes" as required in Step 7(1').
- Dr. Foroosh's proposed combination does not operate in real time as required by claim 14.

160. 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

161. Dr. Foroosh's analysis begins with Salah's disclosure of using a mobile phone to acquire images of a patient's teeth, pointing to its disclosure of providing real-time guidance to a patient in the initial acquisition of those images. EX1002, ¶¶77 (citing exclusively EX1008, 27:21-28:2). Dr. Foroosh next identifies that Salah also discloses analyzing the acquired images using various techniques, including the "not preferred" "deep learning" approach, which can include determining image contours. EX1002, ¶78. Dr. Foroosh concedes that Salah does not disclose that any "deep learning" analysis would be used to guide image acquisition, or that it would even be used to determine a value for an image attribute. EX1002, ¶¶78, 83. Additionally, while Dr. Foroosh refers to Salah's disclosure of "deep learning" when addressing claim 1 step b'), he appears to rely on Maninis to satisfy the "deep learning device" limitation of the challenged claims, only identifying Maninis as allegedly disclosing other limitations associated with the "deep learning device," including that it is "trained by means of a learning base" and the composition of said "learning base." *E.g.*, EX1002, ¶¶107-08.

162. Recognizing that Salah does not expressly disclose determining values for image attributes to guide image acquisition, let alone with a "deep learning

device,” Dr. Foroosh argues that a POSA would be motivated to use the guidance techniques disclosed in Carrier instead. EX1002, ¶81 (alleging that a “POSA would have had reason to use Carrier’s technique for assessing alignment of teeth to guide users in an image acquisition apparatus as taught by Salah”). Dr. Foroosh suggests that Carrier discloses two alignment techniques. The **first** technique is the use of an on-screen overlay that assesses alignment by checking whether identified contours of the patient’s teeth are within a threshold distance of an overlay (referred to herein as the “overlay technique”). EX1002, ¶79. For the **second** technique, Dr. Foroosh reads Carrier’s reference to “ensur[ing] that all required teeth are visible” as indicating that an “image recognition algorithm” should be used to identify which teeth – by number – are present in an image (referred to herein as the “tooth counting technique”). EX1002 ¶¶86, 115, 119-120. As Dr. Foroosh confirmed in his deposition, Dr. Foroosh proposes that the deep learning device would analyze the images in a manner such that both techniques will be applied in his proposed combination. EX2048, 53:4-17.

163. When proposing that Carrier’s alignment techniques are incorporated into Salah’s methods and apparatus, Dr. Foroosh fails to address that Salah already discloses an automatic alignment technique, specifically the alignment of on-screen “references” with “registration marks” placed on the retractors provided with the “acquisition kit.” During his deposition, Dr. Foroosh confirmed he had not relied

on Salah’s “registration mark” alignment technique. EX2048, 178:2-23, 182:7-183:4. I thus understand that this technique is not part of his proposed combination and would thus be displaced by the otherwise redundant techniques incorporated from Carrier.⁵

164. Apparently recognizing that neither Carrier nor Salah disclose the use of a deep learning device trained by means of a learning base for analyzing images in connection with guiding the user to acquire dental images, Dr. Foroosh proposes that “a POSA would be motivated to adapt Salah and Carrier’s image acquisition device to employ a neural network as taught by Maninis[.]” EX1002, ¶86 (alleging that “a POSA would be motivated to adapt Salah and Carrier’s image acquisition device to employ a neural network as taught by Maninis” and discussing the alleged specific functionalities provided by “Maninis’s neural network”). Specifically, Dr. Foroosh proposes that Maninis’s COB neural network would be used to determine semantic contours in images of the patient’s teeth, allegedly providing the contours necessary to apply the “overlay technique” and an identification of each visible tooth to apply the “tooth counting technique.” EX1002, ¶¶84, 87.

⁵ I address both possibilities that the Carrier alignment technique is used in place of and in addition to that in Salah *infra* at Sections VII.A-F.

165. To train the Maninis COB neural network by means of a learning base, Dr. Foroosh alleges that the learning base would comprise more than 10,000 dental images, captured by cell phone, similar in composition to the images that will need to be analyzed when carrying out his Salah/Carrier combination. EX1002, ¶167; EX2048, 103:23-106:5, 222:7-12. Each image would be labeled with pixel-wise contours corresponding to each tooth visible in the image and identified as being associated with the particular tooth (by tooth number) that it borders. *Supra* §§V.A-C.

166. While Dr. Foroosh only discusses these details regarding the learning base in dependent claims 3 and 5, Dr. Foroosh confirmed in his deposition that in his proposed combination, the learning base is trained in the same manner with respect to his analysis of each claim of the '409 patent. EX2048, 103:23-106:5, 222:7-12.

167. Dr. Foroosh indicated that the 10,000 images for the learning base could be captured via several distinct methods, including by practicing the methods of Salah or Carrier *without* Maninis, acquiring them from an existing database, or capturing them manually, and that they could either be labeled manually or by using information from the 3D models generated and updated in the methods disclosed in Salah (as Dr. Foroosh describes in connection with claims 8 and 9). Dr. Foroosh

provides no indication as to which method a POSA would be motivated to use and fails to identify the costs and benefits associated with each.

B. A POSA Would Have Recognized that Dr. Foroosh's Proposed Combination is Inoperable

168. 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 teeth and the tooth they represent), and analyzes images to determine the same. *Supra* §VII.A. However, as also described above, the Maninis COB neural network is incapable of such training and analysis. *Supra* §V.C. 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.

169. Dr. Foroosh appears to fundamentally misunderstand the Maninis COB neural network. In his deposition, Dr. Foroosh opined that 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 network is only

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

170. As described above, 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* §V.C. 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* Figure 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. *Id.*

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

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* §§VII.D-H.

C. A POSA Would Not Have Been Motivated to Combine Carrier’s Alignment Technique with Salah’s Image Acquisition Methods and Apparatus

173. I disagree with Dr. Foroosh that a POSA would have been motivated to combine the alignment techniques disclosed in Carrier with the methods and acquisition apparatuses disclosed in Salah. EX1002, ¶¶81-82 (alleging that a “POSA would have had reason to use Carrier’s technique for assessing alignment of teeth to guide users in an image acquisition apparatus as taught by Salah. . . . [and] Carrier’s techniques for assessing alignment would be used to determine the feedback to Salah’s users when guiding them to acquire dental images”).

174. I note that Dr. Foroosh has identified no motivation to use the Carrier alignment techniques in Salah’s image acquisition process. Dr. Foroosh says only that Salah and Carrier are “directed toward the same field of endeavor, namely dental imaging” and “provide complementary teachings regarding the same subject within that field: namely obtaining dentition images with users’ mobile phones.” EX1002,

¶81. While Dr. Foroosh indicates that “Carrier provid[es] specific details regarding how to effectively assess alignment,” Dr. Foroosh ignores that Salah already provides such details—specifically by teaching the use of “registration marks” to perform an automatic alignment. *Id.*; *supra* §V.A-B.

175. Dr. Foroosh never indicates that Carrier’s alignment process is superior to that already described in Salah, or that Salah’s alignment technique has any shortcomings that a POSA would have reasonably expected to be addressed by Carrier. On the contrary, Dr. Foroosh opines that “[i]n combination, Salah’s and Carrier’s teachings would each perform *substantially the same function* as they provide alone.” EX1002, ¶82. I disagree.

176. As described above, Salah already discloses the use of a particular automatic alignment method in connection with its image acquisition process. Specifically, Salah teaches the use of registration marks, which can be aligned with on-screen references to guide the alignment of dental images with a mobile phone. *Supra* §V.A.

177. A POSA would have recognized that Salah’s alignment technique has distinct advantages in the context of Salah. For example, Salah teaches that the registration marks are placed on retractors, which themselves assist with image capture by exposing the patient’s teeth. *Id.* Dr. Foroosh does not address whether retractors would still be used in his combination when replacing Salah’s alignment

technique with that in Carrier, but a POSA would have understood that the retractors are a useful tool for image acquisition as taught in Salah and should be provided to the patient for use in capturing dental images. *See id.*, §A-B.

178. Salah also teaches that the presence of the registration marks on the retractors in the captured images are important for the determination of the “acquisition conditions” in the later steps in Salah’s method of updating a 3D model. *Id.*, §A. As described above, Salah describes the significant value that comes from the use of these registration marks to the efficiency and accuracy of its later analysis, and a POSA would not have been motivated to remove this advantage from Salah by replacing the use of registration marks with the technique in Carrier. *See id.*, §A-B. A POSA would have understood that the Carrier technique does not provide the same functionality as the registration marks in determining the acquisition conditions because there is no standardized, physical marker present in the image of the patient’s teeth in those techniques from which the acquisition conditions (*e.g.*, angle of the camera in relation to the patient’s mouth) could be determined. *See id.*

179. Furthermore, Salah expressly teaches away from measuring the patient’s teeth to evaluate alignment during the image acquisition process, which is precisely what Carrier requires. EX2048, 162:8-13 (agreeing that Carrier’s overlay technique, which determines the distance of the teeth from an overlay in millimeters, is a “measurement” of those teeth). A POSA would have understood that Salah is

teaching that such a measurement is disadvantageous because the individual teeth are expected to move and may be in unknown relative positions, making them unsuitable as a fixed reference point to ensuring that the target acquisition conditions are met. *Supra* §V.A. A POSA would have further understood that detecting tooth outlines is a more computationally challenging evaluation than comparing a reference to a standardized, geometric symbol placed on the retractors (*i.e.*, “registration mark”), but such a measurement would be unnecessary in view of the proposed use of registration marks and the low level of precision required for alignment in the image acquisition process in Salah. *See id.* As described above, Salah specifically teaches that the advantage of the subsequent analysis it performs to update the 3D model is that a low level of precision is required in acquiring the images in the first instance. *Id.*

180. Accordingly, a POSA would have recognized that replacing Salah’s alignment technique with one that measures the teeth is unnecessary, it would be detrimental to the operation of Salah’s method and apparatus, and it would come at additional computational expense. A POSA would thus not have been motivated to make this change. *See id.*

181. To the extent that Dr. Foroosh is proposing that the alignment techniques in Carrier would be used in addition to Salah’s registration marks, rather than instead of them, Dr. Foroosh has identified no motivation for incorporating an

entirely redundant alignment technique. A POSA also would not have been motivated to add a second alignment technique, particularly one that is expressly disfavored by Salah, to a process that does not require a high level of precision in the first instance. *See id.*, §A-B. This is particularly true in the context of Salah, in which the analysis supporting alignment is carried out by the processor of the mobile phone itself. Mobile phones have limited computing resources and battery life, and a POSA would not have been motivated to add a redundant feature that not only offers no benefit over the existing approach but would also slow down and complicate image acquisition by the patient. *See supra* §V.A.

D. A POSA Would Not Have Been Motivated to Perform Carrier’s Alignment Technique Using Maninis

182. Even if combining Salah and Carrier as Dr. Foroosh proposes, a POSA would not be motivated to perform Carrier’s alignment technique using the COB neural network disclosed in Maninis.

183. Dr. Foroosh concedes that Salah and Carrier do not disclose using a deep learning device trained by means of a learning base in analyzing an image to produce a value for an image attribute. EX1002, ¶83. Dr. Foroosh proposes instead that a POSA would have been motivated to use the “method referred to as ‘Convolutional Oriented Boundaries (COB)’” in the Maninis reference. Specifically, Dr. Foroosh proposes that Maninis’s COB neural network would be used to “divide[] images into regions containing different objects, in both foreground and background,

and then classif[y] the objects into appropriate categories.” *Id.*, ¶83; EX2048, 53:4-17 (confirming that he relies on Maninis as determining “semantic contours” in analyzed images).

184. Dr. Foroosh does not identify any problem or shortcoming with Salah or Carrier’s image acquisition methods that a POSA would have been motivated to try to address by combining one or both references with Maninis. He does not explain how or why a POSA would have been aware of or located Maninis, nor does he explain how any alleged benefit or advantage of Maninis would improve or benefit the methods disclosed in Salah and/or Carrier such that a POSA would have been motivated to combine them with Maninis. I disagree that there is any such problem or shortcoming with Salah or Carrier, or any such benefit or advantage to combining Salah or Carrier with Maninis that would have motivated a POSA to combine the references to arrive at the inventions claimed in the ’409 patent.

185. Dr. Foroosh suggests that Carrier “expressly invites” the incorporation of additional techniques for edge detection and image recognition, and that “Maninis would be a natural selection for such a technique.” EX1002, ¶84. As described further below (*infra* §§VII.E-H.), I disagree that Maninis would be a natural selection for use in Carrier—Maninis is in fact not compatible with the overlay technique disclosed in Carrier. *Supra* §§V.A-B. I also disagree that Carrier “expressly invites” incorporation of untested techniques, particularly the use of a

neural network. Carrier simply says that “any *appropriate* technique (e.g., edge detection image processing, etc.)” may be employed. See EX1006, ¶63; *supra* §V.B. A POSA would not read this general statement as providing a motivation. If anything, because “deep learning” is not “appropriate” for Carrier’s (or Salah’s) alignment technique or apparatus, this statement teaches away from the proposed combination. *Supra* §§V.A-B.

186. I also disagree with Dr. Foroosh that Salah’s reference to “deep learning” as a “not preferred” technique for determining “discriminant information” in an image “would further motivate a POSA to apply Maninis’s teachings to the imaging apparatus of Salah and Carrier.” EX1002, ¶84. On the contrary, Salah is the only cited prior art reference that even refers to “deep learning” in the context of dental images analysis post-acquisition, and Salah teaches that “deep learning” is “not preferred.” *Supra* §V.A. Salah nowhere suggests that “deep learning” can be used during dental image acquisition, especially not with its method that uses a computer program running on a cell phone to provide real time feedback during image acquisition. *Id.* This would be consistent with a POSA’s understanding—explained further below—that Maninis’s COB neural network would be ill-suited to implementation on the processors of mobile phones and as discussed further below, incompatible with the alignment technique disclosed in Salah to **guide** image acquisition.

187. Indeed, even though Salah suggests that “deep learning” “can . . . be used” to determine *discriminant information*, it does not teach a POSA that deep learning can be used *for determining contours* in particular, nor that it would be a “useful option” for doing so. *See supra* §V.A. Furthermore, Salah’s disclosure of “deep learning” is exclusively in the context of analyzing images post-acquisition for the purposes of updating a 3D model.⁶ *Id.* The motivation that Dr. Foroosh allegedly derives from Salah’s disclosure appears to be nothing more than would be understood from the state of the art in image recognition generally, which is that by 2017, neural networks had proven capable of identifying semantic contours in images outside the context of dental imaging.

188. 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, ¶85. 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

⁶ Dr. Foroosh suggested in his deposition that Salah is using the contours from “deep learning” to inform image acquisition. EX2048, 81:20-82:25. As described above, Salah never suggests any such use of “deep learning,” nor does Dr. Foroosh cite any such example in his declaration. *See supra* §V.A.

CNN.” *Id.* Dr. Foroosh does not explain how or why these alleged benefits would have motivated a POSA to include Maninis in the combination of Salah and Carrier, 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 Salah and Carrier in particular. 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.

189. Dr. Foroosh also states that Maninis describes that its technique offers “improvements” over state-of-the-art deep learning techniques for contour detection,” 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. EX1002, ¶85. 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* §V.C. 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 or object recognition. *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.* 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.* Rather, as I explained above, the authors of Maninis determined semantic contours only by using a separate, prior art neural network in parallel with its COB neural network. *Id.* Dr. Foroosh has never suggested that a POSA would have been motivated to use more than one 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.

190. While the use of the Maninis COB neural network and a second CNN for the determination of semantic contours may produce state-of-the-art results with respect to the accuracy of contours, Dr. Foroosh has failed to explain how that improvement in accuracy would benefit the Salah/Carrier combination into which he proposed it would be incorporated. A POSA would have no evidence that further improvements in contour accuracy would translate ultimately into accuracy of image alignment. Regardless, any such improvement in alignment accuracy would be of little if any value to the methods disclosed in Salah and Carrier, each of which expressly required only approximate alignment to achieve their goals, as described above. *Supra* §§V.A-C; *e.g.*, EX1006, [0006]-[0008], [0010], [0060], [0063],

[0011], [0019], [0022], [0034], [0050], [0056], [0059], [0061], [0065]-[0067], [0072], [0084]-[0085], [0089], [0105]; EX1005, [0007]-[0008], [0011], [0105], [0108], [0012], [0020], [0023], [0070], [0095], [0101], [0104], [0106], [0110]-[0112], [0118], [0132]-[0133], [0166], [0186]; EX1008, 27:19-20, 2:21-27, 30:16-21, 31:6-7, 31:12-13, 31:23-24, 31:26-27. The value of any alleged improvement in contour detection would be further minimized if Dr. Foroosh is proposing that Salah's existing registration mark alignment technique is maintained, because a POSA would have recognized that this technique already provides sufficient accuracy for the methods disclosed therein. *See supra* §V.A. Finally, any potential improvement in contour accuracy would be entirely outweighed by the significant costs, downsides, and incompatibility addressed further below.

1. A POSA Would Not Have Used Maninis's Techniques for Carrier's Real Time Alignment Technique for Acquiring Images on a Mobile Phone

191. Significantly, Dr. Foroosh fails to address whether a POSA would have good reason to employ deep learning techniques to perform *real time* analyses on a *mobile phone*. While not required by the '409 patent claim 1, Dr. Foroosh alleges that each claim of the '409 patent is obvious over his proposed combination in which the acquisition device is a mobile phone and the analysis image is analyzed and checked in real time. *Supra* §§V.A-B., VI.B. Dr. Foroosh did not identify any combination in which the acquisition device was not a mobile phone and the analysis

image was not analyzed and checked in real time, nor has he alleged why a POSA would have been motivated to make such a combination or how a POSA would have achieved it. Therefore, I understand it is proper to consider whether a POSA would have been motivated—as Dr. Foroosh alleges—to combine an acquisition device that is a mobile phone and analyzes dental images in real time during image acquisition with the Maninis COB neural network on which Dr. Foroosh relies. EX1002, ¶86 (Dr. Foroosh alleging that “a POSA would be motivated to adapt Salah and Carrier’s image acquisition device [(i.e., a mobile phone)] to employ a neural network as taught by Maninis.”); EX2048, 56:10-15, 68:25-69:8 (real time).

192. Both Salah and Carrier teach that the software performing their respective alignment analyses are run on processors within the mobile devices that are used to acquire the dental images. *Supra* §§V.A-B, VII.B-C. A POSA would have recognized that this disclosure of performing the methods using processors in the mobile device is directly related to the disclosure in Salah and Carrier that these alignment techniques are carried out in real time. *See id.* The need for real time analysis is particularly acute because the Carrier overlay technique that Dr. Foroosh proposes would be incorporated into Salah’s methods requires that the image alignment must be confirmed quickly enough to **guide** the mobile phone operator to capture the image before the mobile phone is moved out of alignment. *See id.*; EX1006, [0054]; EX1005, [0007], [0099]. Where the operator is holding a mobile

phone in their hands, the alignment would likely need to be confirmed and communicated to the user with sub-second timing. *See id.* Otherwise, the relative position of the phone may have moved, either intentionally while attempting alignment or unintentionally due to the unsteadiness of their hands or head, by the time that alignment is confirmed, rendering the overlay alignment ineffective. *See id.* A POSA would have recognized that this is a real risk because, in the case of Salah (which Dr. Foroosh says a POSA would have modified to use Carrier's alignment), the user is an unskilled patient. *Supra* §V.A.

193. That near-immediate guidance is required by Dr. Foroosh's proposed combination is consistent with the messages that Dr. Foroosh identifies in Salah and Carrier as guiding the image acquisition process. Those include on-screen arrows, changing the color of the overlay when alignment is achieved, or using a series of beeps to indicate increasing alignment accuracy. *E.g.*, EX1002, ¶¶47, 77, 129, 130, 132, 136, 220; EX1006, [0063], [0067]-[0070]; EX1005, [0108], [0112], [0114]-[0116]; EX1008, 27:21-28:10. A user cannot be guided as to which direction the device needs to be moved, nor whether they have achieved or are approaching alignment, if the analysis is too slow to provide feedback that is still relevant to the current positioning of the mobile phone at the time the feedback is provided. *See id.*

194. Maninis's COB neural network is a computationally intensive neural network. *See* EX2007 [Shrestha], 2 (confirming, after the '409 patent priority date,

that “[s]uch costly computation complexity makes the deployment of CNN models unaffordable for common PCs and mobile devices”); *supra* §V.C. As described above, Maninis benchmarks the speed of its generic contour detection on what was an expensive, state-of-the-art computer-gaming GPU that is sized to be installed in a large desktop computer. *Supra* §V.C. Even if the mobile phone images were analyzed only by the Maninis COB neural network to determine generic contours—a combination that Dr. Foroosh has not proposed—it would take far longer than the 0.8 seconds disclosed in Maninis Table 3 to run that analysis on a mobile device in July 2017. *Id.* The Titan X GPU employed in the Maninis reference has the ability to process over an order of magnitude more computations per second. *Id.* Even with the conservative assumption that 100% of the “neural engine” processing power on a leading mobile phone in July 2017 could perform the COB technique, such an analysis would take approximately 6 seconds rather than 0.8 seconds to analyze a single image. *Id.* Six seconds is far too long to perform Carrier’s overlay technique because any feedback regarding the position of a phone 6 seconds after the fact would be obsolete as the user would likely be unable to hold the phone or their head perfectly still for such a length of time (let alone for repeated periods of 6 seconds until a matching alignment is identified). *See id.*

195. Furthermore, the above timing only accounts for running the Maninis COB neural network, and not the entirely separate neural network analysis (or the

third process for integrating the neural network outputs) that would be required to determine the semantic contours as Dr. Foroosh has proposed in his combination. *See id.* While Dr. Foroosh has not identified how quickly or efficiently either of the additional neural networks disclosed in Maninis as providing semantic contours would run, even if they ran as quickly as the Maninis COB neural network—which identifies itself as orders of magnitude faster than other comparable techniques—that would at least double the analysis time since the analyses would have to be run in series (because the above assumption regarding the speed of the analysis by the Maninis COB neural network is that 100% of the phone’s processing power is used when running the Maninis COB neural network). *Id.*

196. A POSA practicing Salah and Carrier, both of which only require approximate alignment and which already identify functional alignment techniques for which no deficiencies have been identified, would not have been motivated to employ the Maninis COB neural network alone because it could not achieve the goal of semantic segmentation or object detection identified by Dr. Foroosh. *See supra* §§V.A-C. Equally, a POSA would not have been motivated to employ the Maninis COB neural network along with the other disclosed neural networks necessary to obtain semantic contours, in order to gain the uncertain benefit of improved contour detection when the resulting image acquisition process would require the user to wait more than 11 seconds for feedback on whether the acquired image was in alignment

(assuming conservatively that the second neural network could run as quickly as the Maninis COB Neural network, adding another 5.1 seconds). *See supra* §V.C. Dr. Foroosh's reliance on the speed and efficiency of Maninis as a motivation to combine it with Salah and Carrier is thus entirely lost when Maninis is actually applied in the combination as Dr. Foroosh proposes.

2. A POSA Would Have Recognized the Significant Time and Expense Associated with Implementing Maninis for Carrier's Alignment Technique

197. Dr. Foroosh alleges that a POSA would have been motivated to train the Maninis COB neural networks on more than 10,000 images of patient's teeth that had been captured with mobile phones. EX1002, ¶167; EX2048, 103:23-106:5, 222:7-12. As described above, 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* §V.C. A POSA would thus have recognized that in order to train the COB neural network in Maninis, a POSA would first have needed to obtain such a learning base or capture and label 10,000 images of patient's teeth using a mobile phone. *See id.* But as Dr. Foroosh also recognizes, there is no indication that any such images already existed as of July 2017. EX2048, 109:3-19.

198. Because there is no evidence that a database of mobile-phone dental images had been compiled in the prior art, I understand Dr. Foroosh proposes that a

POSA motivated to combine Carrier and Salah with Maninis would *begin* by practicing Carrier and/or Salah *without* Maninis for as long as it took to acquire more than 10,000 images required under his proposed combination to train Maninis. Dr. Foroosh does not address the time or expense associated with capturing the images using such methods, but I understand from Dr. Kusnoto that it could take even a large dental clinic more than a year to acquire such images. *Supra* §V.C. 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.*

199. Once acquired, a POSA would also need to label these images with pixel-wise contours identifying each individual tooth by number, consistent with the training Dr. Foroosh proposes in his combination. EX1002 ¶¶151; 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. *E.g.*, EX2048, 144:23-145:10 (indicating that he did not have the expertise to identify which teeth are molars in the Carrier Provisional Figures). As I understand from Dr. Kusnoto, accurately assigning tooth numbers in particular is challenging and requires extensive dental anatomy training. *Supra* §V.C. 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).

200. Dr. Foroosh suggests that manual labeling could be done “with assistance of edge detection algorithms (such as a pre-trained version of Maninis’s neural network or using prior edge-detection techniques).” EX1002, ¶198. I understand Dr. Foroosh to be proposing that a POSA could use rough approximations of contours labeled by an algorithm that lacks the benefit of training on dental images (*e.g.*, Maninis’s neural networks as trained on general image sets as disclosed in Maninis).

201. Dr. Foroosh identifies no evidence supporting that this approach was a practice that a POSA would have been motivated to follow in July 2017. Regardless, such a process would likely still require that a dental professional review each rough annotation and could easily be equally or less efficient than performing the annotations manually without this “assistance” once accounting for the accuracy of those annotations and the added complexity of getting those proposals into a program where they can be adjusted manually. *See supra* §V.C.

202. Dr. Foroosh has also proposed using the invention described in the ’409 patent in claim 8, in which 3D models of a patient’s teeth—such as those generated

in Salah—are used in a reverse of the process disclosed in Salah to label the dental images with information from the 3D model, thus forming images that can be added to the learning base. EX1002, ¶197 (“In particular, the zones containing teeth in the reference image could be used (with manual or automatic adjustment as needed) to identify the tooth zones in the updated image.”). Dr. Foroosh cites no disclosure in Salah that suggests using the 3D models or the captured images in this manner. Nor does Dr. Foroosh identify any disclosure in the prior art that would have led a POSA to this solution other than hindsight coming from an awareness of its disclosure in the ’409 patent. *See infra* §VII.F.(addressing Claims 8 and 9).

203. Furthermore, Salah does not disclose that the 3D models or reference images are labeled with the relevant “tooth attributes,” *e.g.*, “tooth number,” such that this information would be available to transfer to updated images to create labeled, historical images. To the contrary, a POSA would have understood that generating a model of a patient’s teeth using the method disclosed in Salah would not require labeling the individual teeth with tooth type or tooth number. Rather, a POSA would have recognized that the delineations between individual teeth are not critical in the alignment of “discriminant information” between an “updated image” and a “reference image” because the “preferred” techniques for identifying “discriminant information” are not capable of determining semantic contours in the first instance (*e.g.*, Canny analysis). *Supra* §§V.A-C.; EX1008, 34:5-35:11.

204. Accordingly, a POSA would not have been motivated to generate historical images for the learning base in this manner and would thus also not have been motivated to practice the Salah/Carrier/Maninis combination in a manner that would meet the limitations of claims 8 and 9 as proposed by Dr. Foroosh.

205. The cost, time, and complexity required to obtain the 10,000 labeled images to practice the combination of Salah, Carrier, and Maninis as Dr. Foroosh proposes would represent a significant disadvantage that could be avoided by using well established image recognition techniques rather than deep learning. *Id.* This disadvantage would far outweigh any marginal, potential benefits in contour detection proposed by Dr. Foroosh. A POSA would thus not have been motivated to combine Salah, Carrier, and Maninis for this additional reason.

E. A POSA Would Not Have Been Motivated to Identify Individual Teeth to Guide Image Acquisition in Either Carrier or Salah

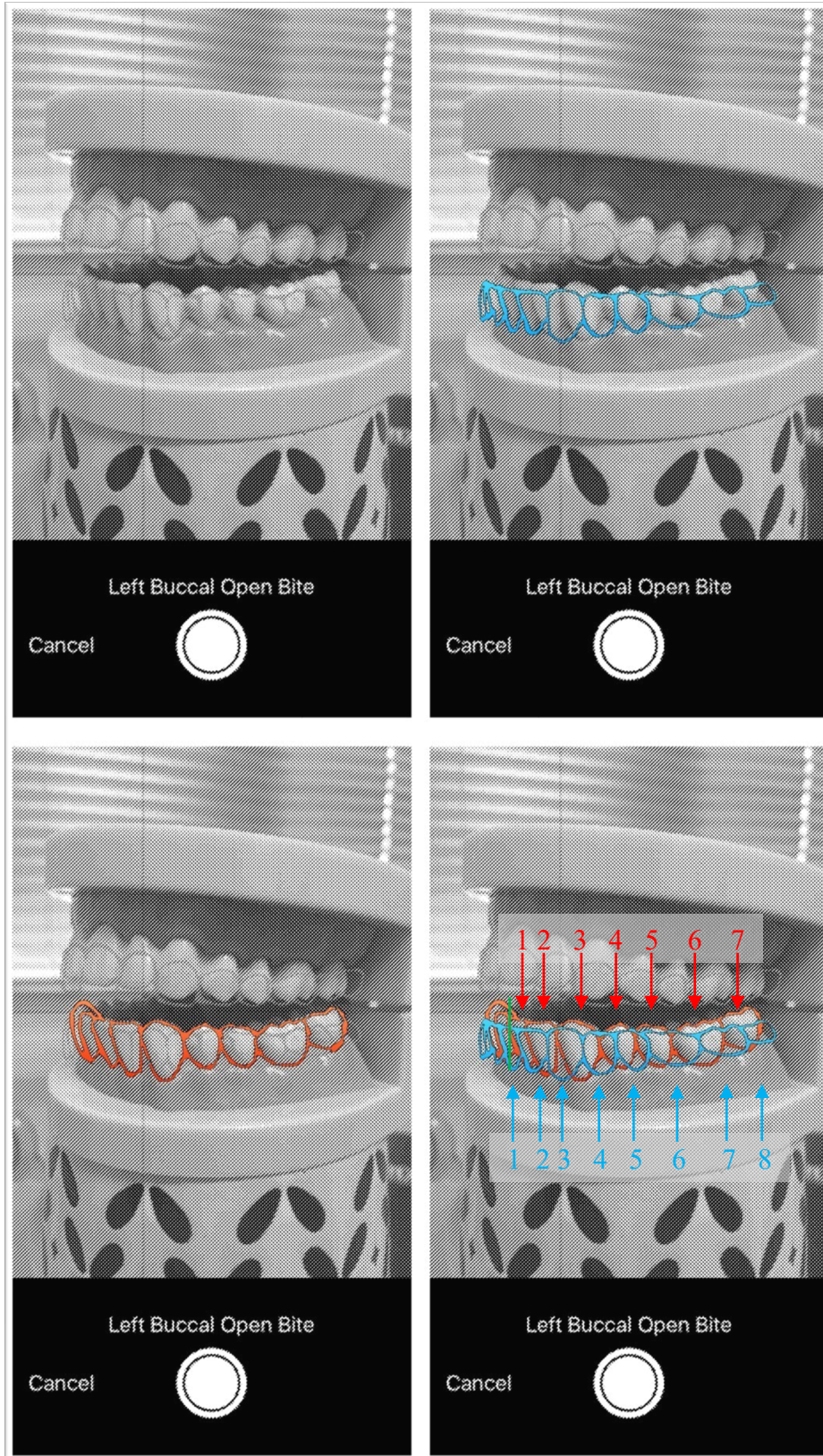
206. Dr. Foroosh proposes a combination in which the Maninis COB neural networks is trained on images labeled with individual contours identified as corresponding to individual teeth by tooth number, and in which the Maninis COB neural network analyzes those images to identify semantic contours of individual teeth in the analysis images. None of the references disclose the recognition of individual teeth, nor would a POSA have been motivated to do so in practicing Dr. Foroosh's proposed combination.

207. As discussed above, Dr. Foroosh recognizes that Carrier does not expressly disclose detecting contours associated with individual identified teeth (semantic contours), and that a POSA would have understood Carrier's overlay technique could be accomplished without doing so. *Supra* §§V.A., VII.B-C. Nonetheless, Dr. Foroosh proposes that a POSA could practice Carrier's alignment technique using such semantic contours. *Supra* §VII.A. Dr. Foroosh identifies no advantage to doing so, and for the reasons discussed above, the Maninis COB neural network cannot identify semantic conducts or perform object identification. *Supra* §§V.A., VII.B-C. Additionally, a POSA would not have been motivated to take on the burden of training and incorporating the Maninis COB and other neural networks for the purpose of recognizing such semantic contours, particularly when there is no identified advantage to doing so in the methods of Salah and Carrier. *See id.* To the contrary, such a change would be detrimental to the operation of Salah, and would render the real time methods in Salah and Carrier unworkable. *See id.*

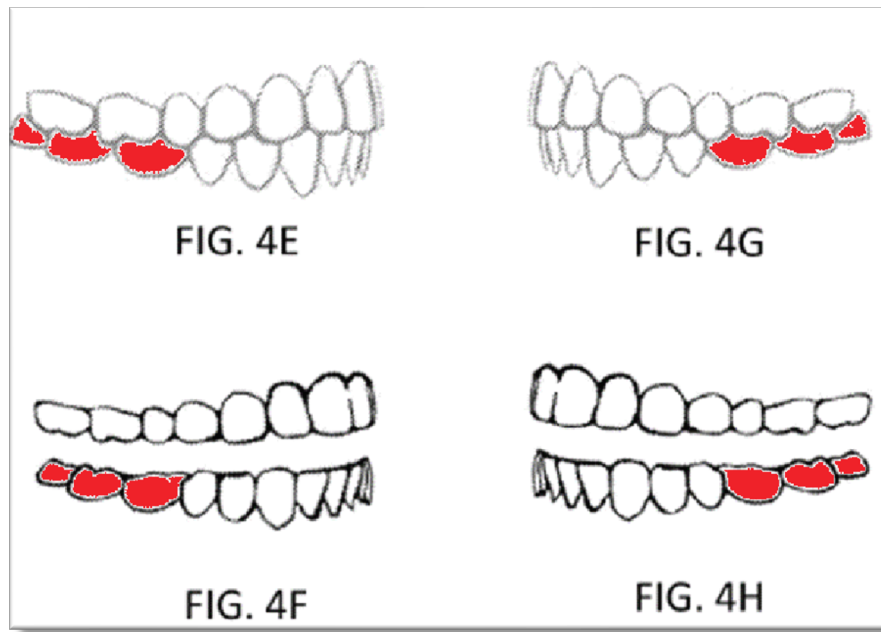
208. Dr. Foroosh also points to Carrier's disclosure "using image recognition algorithms to ensure that all required teeth are visible" as prompting the use of the neural networks in Maninis to identify semantic contours of individual teeth for the purpose of checking the list of present teeth against an alleged list of "required teeth." EX1002 ¶¶80-82, 86, 115, 119-120. However, as explained above, including as informed by my discussion with Dr. Kusnoto, Carrier does not suggest

that an algorithm should be used to check off a list of required teeth as visible in an image. *Supra* §§V.A., VII.B-C. Rather, as explained above, a POSA would have recognized that in the context of acquiring the standard orthodontic images discussed in Carrier, determining a list of present teeth would not be valuable because standard orthodontic images—which are the goal of Carrier—are primarily concerned with the approximate alignment of the “view” rather than with any particular tooth, particularly because patients have varying numbers of teeth. *Id.*

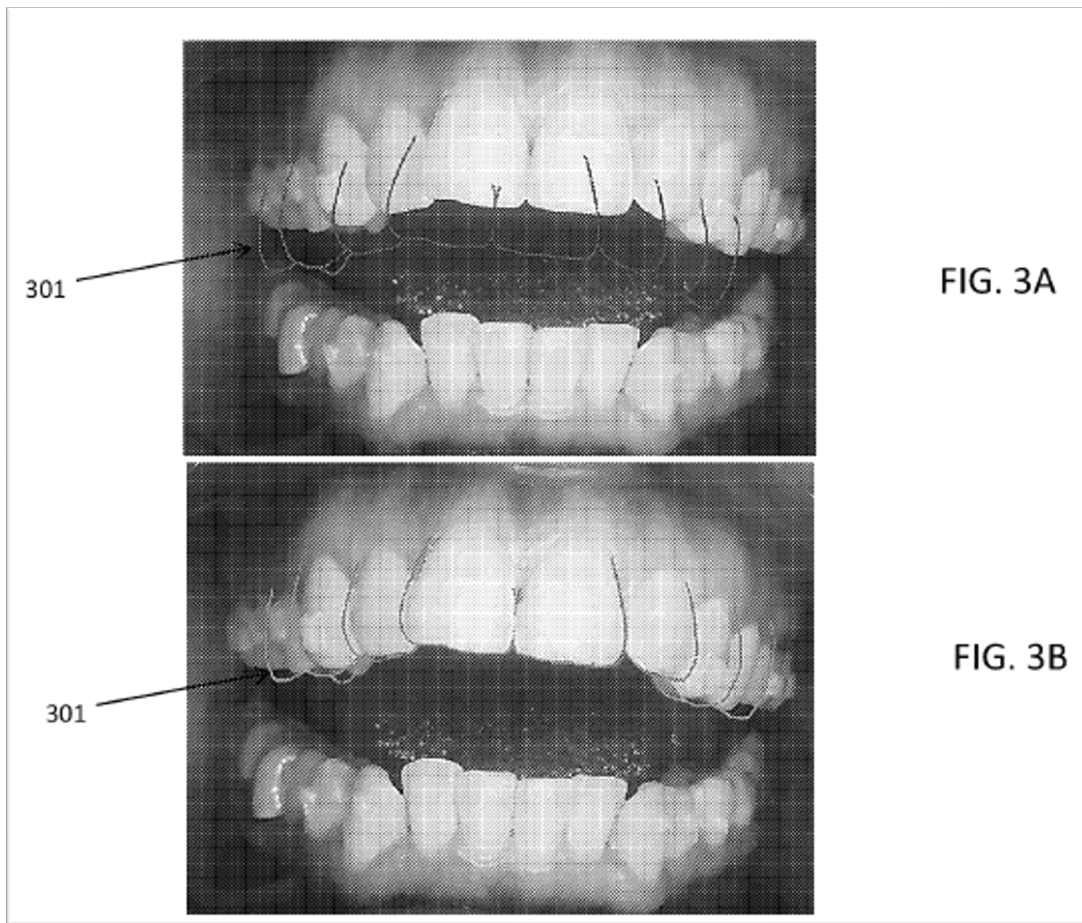
209. Indeed, Carrier already discloses an “image recognition algorithm to ensure that all required teeth are visible” that does not involve checking the list of present teeth against an alleged list of “required teeth”—its overlay technique. For example, Carrier states that “[t]he **overlay** with an outline of teeth in a predetermined view can further provide information such as **required visible teeth** in the predetermined view,” citing Figure 2C and noting that only “an approximate visual match is needed to select the correct distance, correct angle and correct view of the patient’s teeth.” EX1006, [0060]. Even in the cited example in Figure 2C (below), I understand from Dr. Kusnoto that there are only **two** bottom molars in the dental arch while there are **three** bottom molars in the overlay of “required visible teeth”:



EX1006, Fig. 2C (top-left: original; top-right: overlay outlined in blue; bottom-left: modeled teeth outlined in red; bottom-right: modeled teeth counted in red numbers and the corresponding teeth on the overlay counted in blue numbers); *see also* Figs. 4E-4H (depicting three bottom molars in the outlines, as annotated **below** in red).



210. A POSA also would have recognized that unlike checking off a list of teeth visible in an image, the overlay technique guides the approximate distance, angle, and view. *Supra* §§V.A-C., VII. Confirming that a list of expected teeth are visible in the image is unhelpful if those teeth are visible from an inappropriate angle or view. For example, the same teeth are visible in Figures 3A and 3B, but 3A is identified as not being well aligned. *E.g.*, EX1006, Figs. 3A and 3B (below); EX1006, [0063] (Figure 3A is prior to proper alignment, and Fig. 3B is “in alignment with the overlay.”).



211. Accordingly, Carrier would not have motivated a POSA to identify a list of individual teeth present in an image for purposes of “approximate” alignment at all. *Supra* §V.B. Indeed, if a POSA adopted this method using a generic overlay or list that contained a different number of teeth from the patient, the “required teeth” would never be detected and the method would fail. *Id.*

212. Salah also does provide any such motivation. As described above, Salah expressly discourages measuring teeth to guide image alignment, including

because it already discloses a superior technique for doing so that is critical to the 3D modeling process in Salah. *Supra* §V.A.

213. To the extent that Dr. Foroosh is suggesting that a POSA who has already decided to employ Maninis to practice Carrier's alignment techniques in Salah's method would be motivated to determine semantic contours of teeth, that is a purely hindsight driven, circular analysis. Dr. Foroosh identifies no reason a POSA would be motivated to combine Maninis with the Salah/Carrier combination *except* to perform the identified semantic contour analysis, which, as described above, neither Salah nor Carrier teaches in the first instance (and which a POSA could not perform with the Maninis COB neural network and would not be motivated to incorporate by modification for the reasons described above). *Supra* §§V.A-C, VII.

F. A POSA Would Not Have Been Motivated to Create a Learning Base by Altering the Method Disclosed in Salah (Claims 8 and 9)

214. Claims 8 and 9 are each directed to the manner in which the learning base in claim 1 step b') is generated. Specifically, they each require assigning tooth attributes to tooth zones in an updated image by comparing that updated image to a reference model. Dr. Foroosh proposes that a POSA would be motivated to modify the methods disclosed in Salah such that, rather than updating the 3D model with information *from* updated images, a POSA would instead transfer information from the 3D model *to* the updated images and then use those images as the learning base

to train the Maninis COB neural network in the combination of Salah, Carrier, and Maninis. EX1002, ¶¶197-98.

215. Dr. Foroosh does not opine that Salah discloses such a method. The contours Dr. Foroosh identifies are detected as discriminant information in an updated image are generic contours and do not identify individual teeth. EX1002, ¶197 (citing EX1008, 48:9-19, Fig. 6b). Nor is Dr. Foroosh's citation to Salah's disclosure that "regist[r]ation marks" can guide cropping an image to "the elements that were the subject of the initial reference model" relevant to his proposed use of Salah's models as a source of annotations to be applied to the updated images. EX1002, ¶198 (citing EX1008, 33:27-34:2).

216. Furthermore, Dr. Foroosh identifies no motivation for a POSA to modify Salah to achieve what is claimed in the '409 patent. *Supra* §§V.A., VII. As discussed above, Dr. Foroosh has identified no motivation to practice Salah, or the Salah/Carrier combination, *without Maninis* for the extended time it would take to acquire sufficient images to form his learning base. *Supra* §§V.A-C, VII.

217. In particular, Dr. Foroosh identifies no evidence that a POSA in July 2017 would have employed images with pixel-wise contours labeled *by a computer algorithm* as the historical images in a learning base for training any deep learning device. The training bases disclosed in Maninis, on which Maninis reports its results as a cutting-edge technique for contour detection, are database of images with

segmentations that were hand annotated by humans. *Supra* §§ V.A-C, VII. Despite Salah’s disclosure that it produces accurate 3D models, a POSA would have recognized that review by a dental professional to confirm the accuracy of the annotations would likely still be important prior to using them to train any neural network. *Id.* A POSA would have understood that the quality of the training images directly affects the quality of the neural network’s analysis. *Id.* To the extent that it is Dr. Foroosh’s position that a POSA would have been motivated to combine Maninis’s COB neural network because of its potential to provide high-precision contour analyses—despite the express teachings in Salah and Carrier that mere “approximate” alignment is sufficient—a POSA would also have been motivated to employ a dental professional to apply annotations to the training images or at least to validate and adjust each labeled image to confirm that the contours were accurately applied to the training images. *Id.* Furthermore, a POSA would have recognized that the 3D models in Salah require an initial model of each patient’s teeth acquired using physical molding or in-office 3D scanning using specialized equipment. *Id.* Between the time and expense of acquiring 3D models accurate enough to generate annotations, and a dental professional reviewing those proposed annotations, a POSA would not be motivated to create the learning base proposed in Dr. Foroosh’s combination by employing Dr. Foroosh’s proposed modification of Salah. *See id.*

218. Accordingly, I disagree that Dr. Foroosh has demonstrated that a POSA would have been motivated to combine the Salah, Carrier, and Maninis references in a manner that would satisfy the limitations of '409 patent claims 8 and 9.

G. Dr. Foroosh's Semantic Contours Are Not "Image Attributes" (Claim 7)

219. Claim 7 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, 7. Claim 7 requires that each "historical image" in the learning base comprise an "image attribute value" and that "at least one probability relating to *said* image attribute value" is determined during analysis. *Id.* 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 2-6, which embody the "detailed analysis approach" described in the '409 patent, claim 7 embodies the "global analysis" approach. *Supra* §IV.C (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.").

220. As described previously, Dr. Foroosh's proposed combination of Salah, Carrier, and Maninis is one in which the images in the learning base are labeled with

semantic contours (or boundaries)—specifically, with contours identifying the tooth bounded by that contour. EX1002 ¶177 (referring to “section IX.D” “discussing annotations in the context of claim 3”); *id.* ¶151 (describing these semantic contours as satisfying the requirement of “historical tooth zones, to each of which . . . a tooth attribute value is assigned”). Consistently, Dr. Foroosh asserted that the “probabilities” determined by the deep learning device concern the presence of a “set of tooth edges” that “define analysis tooth zone[s]” and “that ha[ve] been determined to contain a particular tooth.” EX1002 ¶¶157-58; *id.* ¶181 (referring to “section IX.D.3”). When addressing Step 7(3’), Dr. Foroosh thus appears to suggest that there are **two values** “related to” the these “probabilities” that satisfy the “said image attribute value” limitation of claim 7: **(1)** “the teeth identified within those boundaries” and **(2)** “the distance between the edges of the tooth and the overlay.” EX1002 ¶181.

221. **(1).** As to **the first**, “the teeth identified within those boundaries” are “*tooth* attributes,” not “*image* attributes,” as Dr. Foroosh confirms when addressing claims 3, 8, and 9. EX1002 ¶¶157-58, 196-200, 210-211. Specifically, the semantic contours constitute “tooth zones” (segments) each assigned a “tooth attribute” (tooth identification). The identification of a tooth associated with a particular segment or “zone” of the image does not “relate[] to the whole image” and is thus not an “image attribute.” EX1002 ¶73 (addressing construction of “image attribute”). This

determination of a tooth attribute cannot satisfy the requirement in Step 7(3') that the deep learning device “determine, for said analysis image, at least one probability relating to said *image attribute value*.”

222. It may be that Dr. Foroosh meant to indicate that a *list* of the teeth that are present in the image, as derived from “tooth attributes” (tooth identifications), is the “image attribute value,” as Dr. Foroosh argues elsewhere. *E.g.*, EX1002 ¶74 (“a list of values for a set of teeth would also be an image attribute, such as a list of all tooth numbers of teeth represented in an image”). Although the '409 patent discloses that such a list can be an “image attribute value” within the meaning of the patent, it *cannot* satisfy the requirements of claim 7, which requires that the historical images in the learning base are labeled with the same “image attribute value.” EX1001, 26:21-23.

223. The neural networks disclosed in Maninis, which determine semantic contours or semantic segmentations, are not capable of being trained on images labeled with *any* “image attribute value,” including a “list of tooth numbers.”

224. Specifically, 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 a “list of tooth numbers.” 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, which indicates whether or not that pixel is part of a contour. *Supra* §§V.C., VII. The Maninis COB neural network is not capable of accepting any inputs other than a vector of values each 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 input representing an image attribute, which would require adding a value to the input vector, the Maninis COB neural network would have no way of mathematically accepting that information, let alone training on it.

225. 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* §§V.C., VII. 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, a tooth number), 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, tooth number), 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 vector training labels 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 a "list of tooth numbers." *Id.*

226. 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. *Id.* 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. *Id.* 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. *Id.*

227. As explained above, the Maninis COB neural network is also trained on a bitmap of values—except that each pixel value is either “0” and “1” because the COB neural network does not recognize object classes—and is thus equally incapable of training on such an “image attribute.” *Supra* §§V.C., VII.

228. As a result, because Dr. Foroosh has proposed that the “deep learning device” is a neural network that determines contours or semantic contours rather than a device that determines wholistic image features (*e.g.*, object recognition), 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” like a “list of tooth numbers” for the image. EX1001 3:26-27; EX1002 ¶73.

229. (2). The **second** apparent “image attribute value” that Dr. Foroosh identifies—“the distance between the edges of the tooth and the overlay”—likewise fails because Dr. Foroosh fails to propose that such an attribute would be included in a historical image in the learning base. Petitioner has thus not identified an “image attribute value” that satisfied both limitations of Step 7(1’) and 7(3’).

230. For the same reason, as with the list of tooth numbers, it is impossible to incorporate such a “distance between the edges of the tooth and the overlay” into the historical images for training any of the neural networks disclosed in Maninis.

231. In other words, claim 7 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 on “tooth attribute values” (pixel-wise annotations). Neural networks that determine semantic contours, as Dr. Foroosh alleges a POSA would be motivated to do, are *incapable* of accepting *any* “image attribute value” (including a list of teeth present in the image). 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 (only tooth attribute values).

232. As a result, Dr. Foroosh has failed to demonstrate that the limitation of Step 7(1’) is satisfied by the proposed combination of Salah, Carrier, and Maninis.

Furthermore, to the extent that Dr. Foroosh intended to suggest that an “image attribute value” would be included in the historical images in his proposed combination, a POSA would not have been motivated to do this 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, as required by claim 7.

H. The Ground References Fail to Disclose that the “Analysis Image is Analyzed and Checked in Real Time” (Claim 14)

233. Dr. Foroosh offers no construction of “real time” in his declaration and indicated in deposition that he has no understanding of what “real time” means within the context of the ’409 patent. EX2048, 69:11-74:12. Accordingly, Dr. Foroosh has failed to offer an opinion regarding whether or how his proposed combination meets this claim limitation.

234. Furthermore, although Dr. Foroosh opines that his proposed combination operates in “real time” as that term is used in Carrier, for the reasons described above, a POSA would have recognized that his proposed combination would *not* be capable of operating in “real time” as that term is used in claim 14. EX2048, [CITE]; *supra* §V.B-C., VII (a POSA would have understood that the Maninis COB neural network alone, or in combination with one of the other two neural networks from other authors disclosed in Maninis, could not operate on a mobile phone to analyze the alignment of images with overlays as described as

occurring in “real time” in Carrier). Accordingly, Dr. Foroosh’s proposed combination does not disclose this claim of the ’409 patent.

235. Furthermore, because Dr. Foroosh’s combination would not operate in real time, a POSA would not have been motivated to combine Salah, Carrier, and Maninis as proposed by Dr. Foroosh to arrive at the invention in claim 14 of the ’409 patent with a reasonable expectation of success.

VIII. CONCLUSION

236. 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 Salah, Carrier, 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 May 28, 2024, in San Francisco, California.

/John Mongan/
Dr. John Mongan