

Filed on behalf of: Align Technology, Inc.

By: Michael T. Rosato (mrosato@wsgr.com)
Matthew A. Argenti (margenti@wsgr.com)
Patrick M. Medley (pmedley@wsgr.com)

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
Patent No. 10,755,409

**PETITION FOR INTER PARTES REVIEW OF
U.S. PATENT NO. 10,755,409**

TABLE OF CONTENTS

	<u>Page</u>
I. Introduction.....	1
A. Background	2
1. Image analysis and computer vision	3
2. Edge detection and contour detection	4
3. Alignment and registration	5
4. Deep learning	6
B. Cited Art	7
1. Salah	7
2. Carrier	8
3. Maninis.....	10
C. The '409 Patent	11
D. Prosecution History	13
E. Level of Skill in the Art.....	14
II. Statement of the Precise Relief Requested for Each Claim Challenged	15
III. Claim Construction.....	15
A. image attribute	15
B. tooth attribute	16
C. Step 1(b')	17
IV. Claims 1-15 are unpatentable as obvious under 35 U.S.C. §103 over Salah, Carrier, and Maninis	18
A. Overview of Ground and Rationale to Combine with Reasonable Expectation of Success	18
B. Claim 1	24
1. Preamble.....	24
2. Step 1(a')	25
3. Step 1(b')	27
4. Step 1(c')	31

5.	Step 1(d')	36
6.	Step 1(e')	37
C.	Claim 2	41
D.	Claim 3	43
1.	Step 3(1)	44
2.	Step 3(2)	45
3.	Step 3(3)	46
4.	Step 3(4)	47
E.	Claim 4	48
F.	Claim 5	48
G.	Claim 6	49
H.	Claim 7	50
1.	Step 7(1')	50
2.	Step 7(2')	51
3.	Step 7(3')	52
I.	Claim 8	52
1.	Step 8(A)	53
2.	Step 8(B)	56
3.	Step 8(C)	57
4.	Steps 8(D)-(F)	59
J.	Claim 9	61
1.	Step 9(A')	62
2.	Step 9(B')	62
3.	Step 9(C')	63
4.	Steps 9(D)-(F)	65
K.	Claim 10	65
L.	Claim 11	66
M.	Claim 12	67
N.	Claim 13	67

O.	Claim 14	68
P.	Claim 15	69
V.	Discretionary Denial is Not Warranted	70
VI.	Grounds for Standing.....	72
VII.	Conclusion	72
VIII.	Mandatory Notices under 37 C.F.R. §42.8.....	73
IX.	Payment of Fees under 37 C.F.R. §§42.15(a) and 42.103	73
X.	Certificate of Compliance.....	74
XI.	Appendix – List of Exhibits.....	75

LISTING OF CHALLENGED CLAIMS

1. A method for acquiring an image of a dental arch of a patient, said method comprising the following steps:

a') activation of an image acquisition apparatus so as to acquire an image, called "analysis image", of said arch;

b') analysis of the analysis image by means of a deep learning device trained by means of a learning base;

c') determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute;

d') comparison of said image attribute value with a setpoint;

e') sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof,

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.

2. The method as claimed in claim 1, in which, in the step b'), at least one analysis tooth zone is identified that represents, at least partially, a tooth on said analysis image, and at least one value of a tooth attribute is determined for said at least one analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute value.

3. The method as claimed in claim 2, in which the step b') comprises the following steps:

1) creation of a learning base comprising more than 1000 images of dental arches, or "historical images", each historical image 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;

2) training of at least one deep learning device by means of the learning base;

3) submission of the analysis image to the deep learning device for it to determine 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,

4) 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.

4. The method as claimed in claim 3, in which said tooth attribute is chosen from a tooth number, a tooth type, a tooth shape parameter, a tooth appearance parameter, a parameter relating to the state of the tooth, an age for the patient, or a combination of these attributes.

5. The method as claimed in claim 3, in which, in the step 1) or 1'), a learning base is created comprising more than 10 000 historical images.

6. The method as claimed in claim 1, in which, in the step b'), all said analysis tooth zones are identified, and at least one tooth attribute value is determined for each analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute values.

7. The method as claimed in claim 1, in which the step b') comprises the following steps:

1') creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image comprising an attribute value for at least one image attribute, or “image attribute value”;

2') training of at least one deep learning device by means of the learning base;

3') 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.

8. The method as claimed in claim 1, in which the learning base is composed according to the following steps:

A) at an “updated” instant, production of a model of a dental arch of a patient, or “updated reference model”, and segmentation of the updated reference model so as to produce, for each tooth, a “tooth model”, and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model;

B) acquisition of at least one image of said arch, or “updated image”, in real acquisition conditions;

C) for each updated image, search for virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called “reference image”, exhibiting a maximum match with the updated image in said virtual acquisition conditions, and acquisition of said reference image;

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

9. The method as claimed in claim 1, in which the learning base is composed according to the following steps:

A') at an initial instant, production of a model of a dental arch of a patient, or “initial reference model”, and segmentation of the initial reference model so as to produce, for each tooth, a “tooth model”, and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model;

B') at an updated instant, acquisition of at least one image of said arch, or “updated image”, in real acquisition conditions;

C') search, by deformation of the initial reference model, for an updated reference model and virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called “reference image”, exhibiting a maximum match with the updated image in said virtual acquisition conditions;

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

10. The method as claimed in claim 1, in which, to create a historical image of the learning base,

identifies one or more “historical” tooth zones on an image, called then assigns, to each identified historical tooth zone, a value for at least one tooth attribute, and/or

assigns to an image a value for at least one tooth attribute.

11. The method as claimed in claim 1, in which the acquisition apparatus is a cellphone.

12. The method as claimed in claim 1, in which the information message is sent by the acquisition apparatus.

13. The method as claimed in claim 1, in which said image attribute relates to

a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image, or a combination thereof, 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, in particular 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 an orthodontic appliance.

14. The method as claimed in claim 1, in which said analysis image is analyzed and checked in real time.

15. The method as claimed in claim 1, in which the deep learning device is a neural network.

Align Technology, Inc., (“Petitioner” or “Align”) requests review of United States Patent No. 10,755,409 to Philippe Salah et al. (“the ’409 patent,” EX1001), currently assigned to Dental Monitoring SAS (“Patent Owner”). This petition demonstrates by a preponderance of the evidence that claims 1-15 of the ’409 patent are unpatentable as obvious.

I. INTRODUCTION

The ’409 patent purports to have invented a method for acquiring images of dental arches involving guiding a user to take images of teeth. However, the co-inventors listed on the ’409 patent (Salah et al.) had published methods for acquiring such images, including most details recited in the claims of the ’409 patent, in a PCT application more than one year before the earliest claimed priority date of the ’409 patent.

The details not expressly found in the listed inventors’ prior publication are found in the relevant prior art. A patent application to Carrier et al. with an earlier effective filing date than the ’409 patent discloses specific details regarding how alignment of an image is computed and used to guide users in acquiring images, including use of edge detection and image recognition that teach key limitations of the ’409 patent’s claims.

While the ’409 patent also recites the use of deep learning at a high level of generality, the prior PCT publication of Salah et al. had already suggested using

such deep learning for tasks including edge detection, and a prior art publication to Maninis et al. had shown how deep learning could be used for edge and object detection on a wide array of image subjects. Thus, as explained in more detail in the petition, the '409 patent fails to patentably distinguish over existing prior art techniques.

Claims 1-15 of the '409 patent should be found unpatentable and cancelled.

A. Background

This petition is supported by the Declaration of Dr. Hassan Foroosh, a Professor in the Department of Computer Science at the University of Central Florida. EX1002, ¶¶1-16; EX1003. As Dr. Foroosh explains and as addressed in further detail below, the involved claims would not have been considered new or non-obvious to a person of ordinary skill in the art ("POSA") at the relevant time.

The '409 patent describes methods involving computer vision and deep learning, with specific focus on analysis of dental images. The brief background described below provides a concise overview of some important elements of computer vision and machine learning that are relevant to understanding what a POSA would have known in these areas at the time of filing of the '409 patent. EX1002, ¶¶29-41.

1. Image analysis and computer vision

Image analysis is often described as the process of extracting information from images that may be subsequently used to perform certain tasks. EX1002, ¶30. Developing methods and computer algorithms to provide such capability has been the subject of study in computer vision in the past few decades. However, it was not until after the early years of last decade, with major breakthroughs in deep learning, that computers managed to also outperform humans in vision tasks. EX1015, 1-2, 8; EX1002, ¶31. Hardware developments 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.

Computer vision, in simple terms, is the science of making computers see. Vision tasks can be divided into three levels: low-level, mid-level, and high-level tasks. EX1016, 1538-39, 1541; EX1002 ¶32. Accordingly, computer vision algorithms were traditionally developed to handle low-level, mid-level, and high-level vision problems. Low-level computer vision methods process images at pixel-level (e.g., quantization, encoding, image restoration, image enhancement); mid-level computer vision methods produce feature-level representations of images (e.g., color features, edge features, 3D geometric features); and high-level computer vision aims to understand image contents and contextual relationships (e.g., object classification, localization, and instance recognition). EX1002, ¶¶32-

33. Boundary-based representations, such as edges and contours defining separation between regions in images, are common mid-level features. *Id.*

The development of deep learning led to a new class of computer vision methods that learn features from data, as well as taking advantage of high-level labels to learn better features, while still relying also on low-level image properties to extract features. These allow a device to receive raw image data and directly learn to accomplish a vision task. EX1002, ¶34.

2. Edge detection and contour detection

Edges, contours, boundaries, and discontinuities refer to similar concepts involving a representation of a discontinuity in an image. Edges and contours have been used for various computer vision, from high-level (e.g., classification, object recognition, annotation) to low-level (e.g., registration, alignment) tasks. EX1002, ¶35.

Due to the key role of edges in computer vision, researchers developed deep-learning based methods for edge and contour detection. EX1017, 1-2; EX1007, Abstract; EX1002, ¶36. Instead of binary labels (1 and 0) for edge locations, deep-learning methods lead to multiple labels in the contour image. EX1017, 5-9; EX1007, 8, 9-12; EX1002, ¶36. Each contour is labeled based on high-level labels (e.g. class, instance, etc.). Thus, deep-learning methods allow categorization of different objects in an image into specific classes. EX1002, ¶37. These type of

contours or boundaries are sometimes referred to as “category-specific” or “semantic.” *Id.*; EX1017, 9; EX1007, 11. Thus, a deep learning device trained to look for features based on a training set with high-level labels can return features (e.g., edges) specific to those labels. For instance, a machine trained to look for the contours of different types of teeth in intraoral images would detect edges as well as label edges and the regions they define based on which type of tooth they belong to. EX1002, ¶37.

3. Alignment and registration

Alignment and registration refer to a processing step in computer vision allowing comparison of two or more data representations, e.g., two images, a 3D object and an image, etc. EX1002, ¶38. These methods can look for a mapping function minimizing dissimilarity or maximizing similarity related to a label. *Id.* Image labels can indicate presence/absence of an object or class, object location, or an alignment score with a template or with a viewpoint of a 3D object. EX1018, Abstract, 5-7; EX1019, 391-92, 396, Fig. 3 (example images for “tooth dataset”); EX1002, ¶38. Prior to 2017, techniques had been developed to propagate annotations between similar 2D images (such as between video frames), as well as to project annotations from a 3D object to a corresponding region of a 2D image representing the object. *See* EX1013, 3689-90; EX1014, 3198, 3200-01; EX1002, ¶38.

4. Deep learning

Deep learning refers to machine learning methods relying on many feature-learning layers to model the often-non-linear relationship between inputs (the training set) and outputs (probabilities of classes or instances). These models are referred to as “deep” because they have many “hidden” layers between input and output, with these layers potentially comprising millions of parameters to be adjusted through a training process using a training dataset, to produce accurate outputs for unseen test data. EX1002, ¶39.

By 2017, the best results in computer vision arose from learning models based on deep neural networks (DNNs), such as deep Convolutional Neural Networks (CNNs). EX1007, 2. For DNNs, hidden layers’ output may be viewed as a feature map, with features learned through training. EX1002, ¶¶40-41. Earlier layers represent lower-level features, whereas deeper layers correspond to higher-level features. *Id.*; see EX1007, 3 (fine-scale features in early layers, large-scale features in later layers). In image classification or object recognition applications, the output feature map of the first layer corresponds almost to an edge map, while deeper feature maps correspond to higher-level features, such as object parts, shapes, and class. *Id.* By July 2017, DNNs had proven to be remarkably successful in computer vision and surpass human performance, due to their ability to extract multiple levels of features optimized by training with large training datasets.

EX1002, ¶41; EX1007, 7-9 (“human” and “super-human” performance in various tasks).

B. Cited Art

1. Salah

PCT Publication No. WO2016066651 to Salah et al. (“Salah”) is entitled “Method for Monitoring Dentition.” Salah names four inventors, matching four of the five listed on the ’409 patent; however, Salah was published on May 6, 2016, more than 1 year before the earliest claimed priority date of the ’409 patent. Salah is prior art under 35 U.S.C. §102(a)(1). As Salah was published in French (EX1012), a certified translation has been submitted as EX1008.

Salah discloses methods for monitoring patients’ teeth, such as during orthodontic treatment. EX1008, Abstract, 1:12-22, 8:10-18. Salah teaches methods of obtaining 2D images of patient teeth, comparing those teeth to a 3D model, and using those comparisons to update the model and measure treatment progress. EX1008, 1:26-2:19, 46:15-19, Fig. 1. Salah’s analysis can include using contour detection on an acquired image to aid mapping to a reference model, and Salah discloses that this contour detection can be performed using deep learning.

EX1008, 34:11-14, 36:14.¹ These 2D images may be obtained by a patient or relative using a device such as a cell phone, which may also perform the evaluation process. EX1008, 2:21-27, 31:4-7. Salah also discloses that the image-acquisition apparatus can guide the user to obtain appropriate images, including giving feedback on image acceptability and camera positioning. EX1008, 27:21-28:10; EX1002, ¶¶45-46.

2. Carrier

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 Application No. 16/827,594 (EX1010), filed March 23, 2020, and Application No. 15/803,718 (EX1011), filed November 3, 2017, as well as to Provisional Patent Application No. 62/417,985 (EX1006), which was filed November 4, 2016. Each application in Carrier’s priority chain was copending with the previous application and claimed priority to each earlier application in the chain, and each non-provisional application shares a common specification.

¹ Salah uses the term “deeplerning,” which a POSA would have recognized to be a typographical error meant to refer to “deep learning.” *Id.*; EX1002, ¶101.

The '985 provisional provides §112 support for at least one claim of the '923 publication. The chart below illustrates exemplary support in the '985 provisional for claim 1 of the '923 publication. *See* EX1002, ¶¶48-61.

US20210068923	US62/417,985
1. A computer-implemented method comprising:	EX1006, ¶18.
providing a mobile application on a mobile device associated with a patient, the mobile application configuring a user interface for display on the mobile device;	EX1006, ¶¶54-55, 73-75, 94-95, Figs. 5A, 8.
guiding the patient through the user interface to take a first plurality of photos of teeth of the patient, the first plurality of photos corresponding to a first plurality of predetermined views of the teeth, wherein guiding the patient to take the first plurality of photos comprises showing a first plurality of overlays over the teeth in accordance with the first plurality of predetermined views;	EX1006, ¶¶36, 61, 73-75, 85, Figs. 4A-5A, 8, 10.
guiding the patient through the user interface to take a second plurality of photos of the teeth, the second plurality of photos corresponding to a second plurality of predetermined views of their teeth, wherein guiding the patient to take the second plurality of photos comprises showing a second plurality of overlays over the teeth in accordance with the second plurality of predetermined views; and	EX1006, ¶¶36, 61, 73-75, 85, Figs. 4A-5A, 8, 10.
transmitting the first plurality of photos and the second plurality of photos so the first plurality of photos and the second	EX1006, ¶86, Fig. 8.

US20210068923	US62/417,985
plurality of photos can be evaluated for medical treatment.	

This petition provides parallel citations to both the '985 provisional and to corresponding disclosure in the '923 publication, demonstrating that all matter relied on is found in both documents. Accordingly, under *Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375 (Fed. Cir. 2015), Carrier is available as prior art under 35 U.S.C. §102(a)(2) as of November 4, 2016.

Carrier teaches methods to obtain images of patients' teeth for therapeutic use, including images obtained by a mobile phone. EX1006, Abstract, ¶7; EX1005, Abstract, ¶8. The device provides guidance to users to help take high-quality images, including providing an overlay such as an image or outline of teeth that a user may use to align the imaging apparatus. EX1006, ¶¶54-55; EX1005, ¶¶99-100. The device can assess alignment by measuring distance between teeth and the overlay by performing edge detection on teeth and perform image recognition to ensure required teeth are visible. EX1006, ¶¶63, 81; EX1005, ¶¶108, 128; EX1002, ¶49.

3. Maninis

Kevis-Kokitsi Maninis et al. published "Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks" ("Maninis," EX1007) no later than May 2, 2017, on arXiv.org. See EX1002, ¶¶62-67; EX1007, 1; EX1022

(arXiv.org publication policies); EX1023 (document unchanged); EX1026 (same); EX1024 (dissemination by authors); EX1025 (direct download link); EX1021 (earlier version with same material disclosure). Maninis is prior art under 35 U.S.C. §102(a)(1).

Maninis discloses methods for image classification and contour detection using a Convolutional Neural Network (CNN). EX1007, Abstract. Maninis teaches a “deep learning architecture” that produces significant improvements over state-of-the-art techniques. EX1007, 3, 8-9. Maninis uses a neural network trained on precisely annotated images to detect contours defining regions containing objects based on detected probabilities for contour locations. EX1007, 3-4, 5-6, 13, Fig. 2. Maninis also demonstrates the ability to label the specific object contained within each identified region, again improving over the state of the art. EX1007, 11-12, Tables 4-7, Fig. 16; EX1002, ¶68.

C. The '409 Patent

The '409 patent is entitled “Method for Analyzing an Image of a Dental Arch.” EX1001, cover. The '409 patent describes acquiring dental images using an image acquisition apparatus, which may be a cellphone or various other types of widely-used devices. EX1001, Abstract, 8:56-63. Claim 1, the only independent claim, recites steps of acquiring an image, analyzing the image with a trained deep learning device, determining a value for an image attribute based on the analysis,

comparing the value to a setpoint, and sending an information message based on the comparison. EX1001, 32:13-33. The information message must be related to one or more of several categories: “the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof.” *Id.* If the setpoint is not respected, the apparatus guides the user to acquire a new image. *Id.*; EX1002, ¶42.

The '409 patent admits that many of the steps of its method are already found in the prior art. For example, the '409 patent repeatedly admits that Salah teaches many of the steps recited in its claims. *See* EX1001, 2:42-47, 3:4-10, 7:9-33, 7:50-54, 8:55-56, 9:36-39, 14:48-50, 17:65-18:24, 19:35-37, 20:48-53, 22:7-11, 23:41-24:6, 25:21-28, 27:16-32 (repeatedly acknowledging that WO2016066651 (Salah) teaches techniques described in the specification); EX1002, ¶¶43-44. *Compare* EX1001, Figs. 10-12D, 16, *with* EX1008, Figs. 1-2, 4a-d, 5c-6b (same figure images and process).

Regarding the remaining claim limitations, these elements are also found in the prior art as discussed in detail in §IV below.

D. Prosecution History

The '409 patent was filed as U.S. Application No. 16/031,201 (“the '201 application”) on July 10, 2018, claiming priority to French Patent Application No. 1756947, July 21, 2017.

In the only Office action, original claims 1-10 were rejected as anticipated by U.S. Patent Publication No. 2015/0132708 to Kuo et al., and claims 11-13 were rejected as obvious over Kuo in view of U.S. Publication No. 2017/0049311 to Borovinskih. EX1004, 808-31. The claims were also rejected under §112(a) for lack of enablement and §112(b) for indefiniteness. EX1004, 805-08.

Patent Owner responded by amending claim 1 to recite its last limitation, “to check whether the analysis image respects the setpoint...” and by arguing that Kuo neither disclosed this limitation nor disclosed two other limitations: the “image attribute” recited in steps c’ and d’ and the sending of an information message recited in step e’. EX1004, 853-60. Patent Owner argued that an “image attribute...relates to a whole image,” and pointed to the language found at EX1001, 3:23-4:32, and in claim 13 of the '409 patent as exemplifying such image attributes. EX1004, 858-60. Patent Owner further argued that the information message of step e’ must “relate to an acquisition of an image,” and in particular must relate to one of the categories recited in that step. *Id.*

The Examiner responded by allowing the application. EX1004, 868-69. In the reasons for allowance, the Examiner noted that Kuo “does not teach or suggest, among other things, ‘comparison of said image attribute value with a setpoint; sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired, 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.’” *Id.* The Examiner did not comment on whether Kuo or other art of record taught any of the other alternative message subjects recited in claim 1, such as being related “to the position of the acquisition apparatus in relation to said arch ... or to the opening of the mouth.” *See id.*

E. Level of Skill in the Art

As Dr. Foroosh explains, a person of ordinary skill in the art (POSA) by July 21, 2017, would typically have an advanced degree, such as a Masters’ Degree or Ph.D. in computer science, computer engineering, electrical engineering, 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. See EX1002, ¶¶69-70.

II. STATEMENT OF THE PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED

Petitioners request review of claims 1-15 of the '409 patent under 35 U.S.C. §311 and AIA §6. The grounds for relief are as follows:

Ground	Claims	Description
1	1-15	Unpatentable under 35 U.S.C. §103 as obvious over Salah in view of Carrier and Maninis

III. CLAIM CONSTRUCTION

Claim terms should be given their ordinary and customary meaning as would be understood by a POSA in light of the specification. 37 CFR §42.100(b); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312-13 (Fed. Cir. 2005) (en banc). Unless otherwise stated, this petition applies the ordinary and customary meaning of the claim terms. The following terms warrant discussion.

A. image attribute

The '409 patent describes an image attribute as an attribute that has a value that “relates to the whole image and not to part of the image.” EX1001, 3:23-37.

The '409 patent lists examples of what an image attribute may 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.

EX1001, 3:38-49; *see also id.*, claim 13. Applicant also argued this meaning of image attribute during prosecution. EX1004, 858-59. An image attribute value may also be a list of such values for a set of teeth, such as a list of all tooth numbers represented. EX1001, 26:16-26. In light of this description, the term “image attribute” includes an attribute having a value (which may be a set of values for multiple teeth) relating to an image as a whole, including an attribute related to any one or more of the examples listed above. EX1002, ¶¶73-74.

B. tooth attribute

The ’409 patent distinguishes image and tooth attributes, with image attributes relating to an image as a whole and tooth attribute relating to a tooth. *See* EX1001, 1:22-28, 3:23-37. A tooth attribute may nevertheless define an image attribute. EX1001, 3:59-62. For example, a tooth attribute can be a mean separation of a single tooth from an appliance, and an image attribute can be a mean separation of all teeth from the appliance. EX1001, 28:58-65. The ’409

patent also lists examples of tooth attribute categories. *See* EX1001, 8:7-50. Claim 4 recites some of these categories, such as tooth type, tooth number, and tooth shape parameter. Accordingly, the term “tooth attribute” includes any attribute related to a tooth. EX1002, ¶¶75-76.

C. Step 1(b')

Step 1(b') recites “analysis of the analysis image by means of a deep learning device trained by means of a learning base.” Align intends to argue in district court that this is a means-plus-function limitation for which the specification does not provide sufficient structure to discern the claim’s full metes and bounds. Even if the Board agrees, that does not preclude an analysis that the claims fail to distinguish from the prior art.

Based on Patent Owner’s assertion of this claim against Align, Align understands Patent Owner does not view the term as indefinite. To date, Patent Owner has not identified whether it interprets this claim element as a means-plus-function element.² However, regardless of whether step 1(b') is construed as a

² As discussed below in §V, claim construction briefing has not yet been scheduled in district court. Align will keep the Board apprised of any constructions for this term proposed by the parties or adopted by the district court.

means-plus-function limitation, neither the claim language nor the relevant disclosure in the specification distinguishes over the prior art, as explained below for claim 1. For example, the specification repeatedly provides a generic description of “the deep learning device [as] preferably a neural network,” which is met by the deep learning neural network of Maninis. EX1001, 1:24-25; *see* EX1007, 2-3.

For the purposes of this petition, Petitioner identifies the recited function as analyzing an image and the structure to encompass a device comprising a deep learning neural network trained with a learning base. *See* EX1001, 1:24-25, 1:39-40, 1:51-2:3, 3:14-18, 4:5-6, 8:41-50, 13:64-65, 24:14-15, 24:46-47, 24:66-25:4, 28:27-29, 29:24-25, 29:54-55; *see also id.*, 16:13-50 (non-limiting examples of neural networks).

IV. CLAIMS 1-15 ARE UNPATENTABLE AS OBVIOUS UNDER 35 U.S.C. §103 OVER SALAH, CARRIER, AND MANINIS

A. Overview of Ground and Rationale to Combine with Reasonable Expectation of Success

As discussed in detail below, Salah in view of Carrier and Maninis renders claims 1-15 obvious.

Salah discloses methods useful for monitoring patient’s dentitions, including by acquiring images of the patient’s teeth at various times during orthodontic treatment. EX1008, Abstract, 8:19-28. The images can include two-dimensional

images, which may be acquired by an acquisition apparatus such as a mobile phone. EX1008, 13:24-28, 31:4-9. Salah describes that such two-dimensional images can be obtained by users without training in dental arts. *Id.* To aid such users, Salah teaches that the image acquisition apparatus may assess whether a photo is acceptable and assist in positioning the apparatus. EX1008, 27:21-28:2. For example, the device may include marks on a viewfinder to guide alignment, and the device may provide feedback indicating how to adjust alignment. *Id.*; EX1002, ¶77.

Salah discloses that analysis of acquired images may include determining contours, which may involve processes such as edge detection, object detection, and deep learning. EX1008, 34:5-36:14. To the extent that Salah does not expressly describe alignment and acceptability guidance would necessarily involve these analysis steps, or generating values to compare to a setpoint, these aspects would have been obvious to a POSA and are disclosed by Carrier. EX1002, ¶78.

Carrier discloses methods for guiding users in acquiring dentition images with an image acquisition apparatus such as a mobile phone. EX1006, ¶7; EX1006, ¶9. Similar to Salah, Carrier teaches methods for “increasing the accuracy of the alignment of the camera lenses with teeth of the patient” to produce “[h]igh quality dental images.” EX1006, ¶57; EX1005, ¶102. Carrier provides an overlay (e.g., an outline of example teeth) to guide the user to align the image acquisition apparatus

(e.g., cellphone) to capture an image of a patient's teeth. EX1006, ¶¶57, 63, Fig. 6B (below); EX1005, ¶¶102, 108, Fig. 6B; *cf.* EX1008, 30:16-21 (using augmented reality to superimpose semitransparent image of reference model for apparatus alignment).

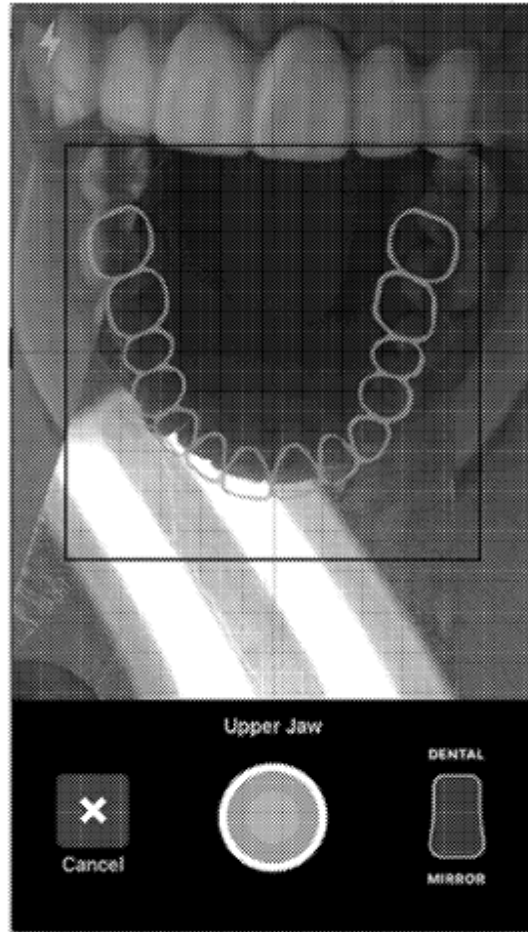


FIG. 6B

To assess alignment, Carrier teaches detecting whether all required teeth are visible and whether the imaged teeth are within a threshold distance of the overlay.

EX1006, ¶¶63, 81; EX1005, ¶¶108, 128; EX1002, ¶¶79-80.

A POSA would have had reason to use Carrier's technique for assessing alignment of teeth to guide users of an image acquisition apparatus as taught by Salah. Both references are directed toward the same field of endeavor, namely dental imaging. Both references provide complementary teachings regarding the same subject within that field: obtaining dentition images with users' mobile phones. Both references teach assessing alignment and using the assessment to guide users, with Carrier providing specific details regarding an effective manner of assessing alignment based on the appearance of the patient's teeth to ensure high-quality dentition images. A POSA would thus have considered Carrier's teachings directly applicable to supplement to Salah's imaging methods. EX1002, ¶81.

The resulting combination would use each reference's teachings to perform substantially the same function as they provide alone, with Carrier's techniques for assessing alignment used to determine the feedback to users when guiding them to acquire dental images as taught by Salah. POSAs would thus have had a reasonable expectation of success, as the combination would be one of complementary techniques, applied in understood ways, to produce a predictable outcome. EX1002, ¶82.

To the extent that Salah and Carrier do not expressly teach using a deep learning device trained by means of a learning base for image analysis, this would

have been obvious to a POSA and is taught by Maninis. Maninis teaches a method referred to as “Convolutional Oriented Boundaries (COB),” which employs a deep learning device to perform image classification. EX1007, Abstract. Maninis provides methods of “contour detection” using “Convolutional Neural Networks (CNNs).” *Id.* Images are divided into regions containing different objects, which are classified into appropriate categories. *See* EX1007, 1-2, 5-6, 8, 11-12; EX1002, ¶83.

Although Maninis is not restricted to dental imaging, Maninis describes broadly applicable techniques for image classification and contour detection using a deep learning device trained using a learning base. EX1002, ¶84. Carrier expressly invites the use of such techniques, teaching that its edge detection be achieved using “any appropriate technique (e.g., edge detection image processing, etc.)” and that detecting presence of required teeth may be performed by “image recognition algorithms.” EX1006, ¶¶63, 81; EX1005, ¶¶108, 128. Similarly, Salah teaches contour detection using any of a variety of known methods, including object detection and deep learning, which Salah describes as optional, though not necessarily preferred. EX1008, 34:5-36:14. Salah’s identification of some methods of image analysis as more preferred would not have dissuaded a POSA from using deep learning, as Salah expressly identifies deep learning as an appropriate option,

and as discussed below, image analysis trends favored greater adoption of deep learning. EX1002, ¶84.

Maninis motivates employing deep learning techniques for object and contour detection. Maninis teaches that the “latest wave of contour detectors takes advantage of deep learning to obtain state-of-the-art results.” EX1007, 2. Maninis provides a flexible, “modular” technique and “can incorporate and benefit from future improvements in the base CNN.” *Id.* Maninis also achieves improvements relative to “recent approaches to contour detection using deep learning.” EX1007, 3, 9 (“Overall, comparing to previous state-of-the-art, we get a significant improvement at a fraction of the computation time (24.37 to 0.79 seconds).”), 11-12 (improved object identification); EX1002, ¶85.

Based on Maninis’s flexibility and prospect to improve performance, a POSA would have been motivated to analyze images using a deep learning device as taught by Maninis to improve an image acquisition apparatus as taught by Salah and Carrier. For example, a POSA would have been motivated to adapt Salah and Carrier’s image acquisition device to employ a deep learning architecture as taught by Maninis, trained on images relevant to Salah’s and Carrier’s dental imaging applications. Maninis’s deep learning architecture would provide specific functionalities required by Salah and Carrier, including the ability to identify objects (such as the required teeth Carrier proposes using image recognition

software to identify) and to detect contours (such as the tooth edges Carrier uses to assess alignments), to aid in guiding image acquisition, as taught by Salah.

EX1002, ¶86.

A POSA would have reasonably expected such a combination to work predictably and successfully, as Maninis teaches that its technique is modular, compatible with multiple different architectures, and yields improvements in both speed and accuracy over prior state-of-the-art techniques. EX1007, 8-13. Adapting Maninis’s techniques to an image acquisition apparatus as taught by Salah and Carrier would also be consistent with trends in computer vision; indeed, Maninis notes that the use of deep learning constituted the trend for the “latest wave of contour detectors,” allowing them “to obtain state-of-the-art results.” EX1007, 2; EX1002, ¶¶87-88; *supra* §I.A.

The discussion of claims 1-15 below provides further detail regarding POSAs’ rationale to combine Salah, Carrier, and Maninis with a reasonable expectation of success.

B. Claim 1

Salah, Carrier, and Maninis teach each element of claim 1, as discussed below. EX1002, ¶¶89-141.

1. Preamble

A method for acquiring an image of a dental arch of a patient, said method comprising the following steps:

To the extent the preamble is limiting, Salah and Carrier teach such a method. Salah teaches acquiring images of patients' dental arches. EX1008, Abstract, 12:20-24. Carrier likewise teaches methods of obtaining images of patients' teeth "for therapeutic use." EX1006, ¶7; EX1005, ¶9. For example, both references teach using a mobile device that guides users to acquire dental images. *See, e.g.*, EX1008, 27:21-28:2, Figs. 5A-D; EX1006, ¶¶59, 96, Figs. 2A-3B, 5C-7C; EX1005, ¶¶104, 173, Figs. 2A-3B, 5C-7C; EX1002, ¶¶89-93.

2. Step 1(a')

a') activation of an image acquisition apparatus so as to acquire an image, called "analysis image", of said arch;

Salah and Carrier teach step 1(a'). For example, Salah teaches that a user (such as a patient or relative) may acquire dentition images using an image acquisition apparatus such as a cellphone. EX1008, 31:4-7. This apparatus can have a camera or video that includes a viewfinder or screen to aid in alignment. EX1008, 27:11-13. The use of such an apparatus to acquire images displayed on a viewfinder or screen when aligning the apparatus constitutes the recited "activation" to acquire an "analysis image." EX1002, ¶94.

Carrier provides similar disclosure. Carrier teaches acquiring images of a patient's teeth (including the patient's dental arch) using a "mobile device such as a smartphone." EX1006, ¶94; EX1005, ¶171. This includes "viewing the patient's teeth, for example, on a screen of a mobile phone" and "displaying, on the screen,

an overlay comprising an outline of teeth in a predetermined view, wherein the overlay is displayed atop the view of the patient's teeth." EX1006, ¶7; see EX1005, ¶8; *cf.* EX1001, 25:66-26:8 (analysis image may be photo or video, including real-time image appearing on screen); EX1002, ¶95.

Carrier teaches that an overlay can be provided to aid alignment. EX1006, ¶59; EX1005, ¶104. Carrier illustrates in figures several examples of such mobile phone images, including an overlay superimposed on an image of a patient's dental arch acquired during alignment. Figs. 3A-B, reproduced below, show two such images, with the camera being moved between Figs. 3A and 3B to align overlay 301 with the patient's upper arch. *See* EX1006, ¶¶62-63; EX1005 ¶¶107-108.

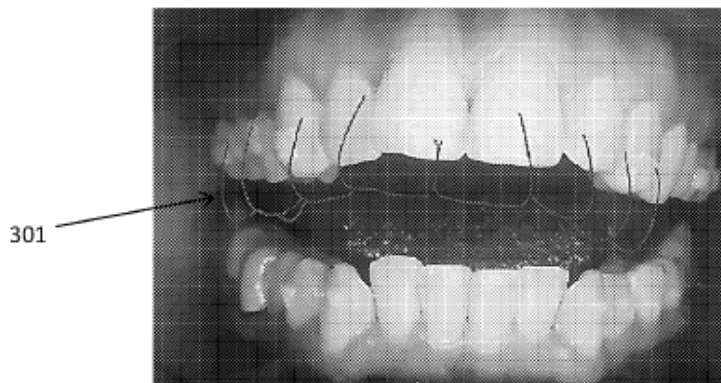


FIG. 3A

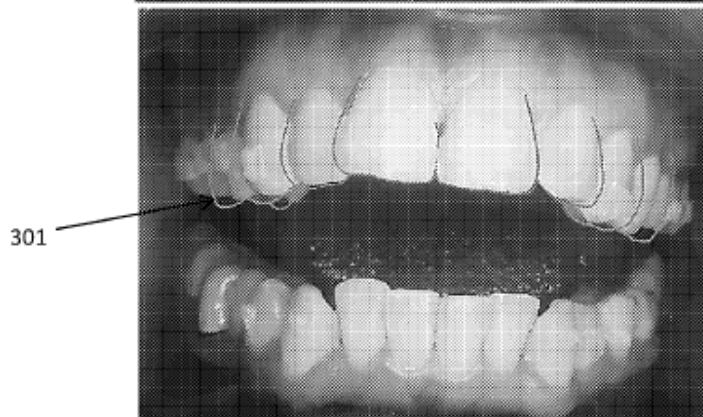


FIG. 3B

EX1006, Figs. 3A-B; *see* EX1005, Figs. 3A-B; *see also, e.g.*, EX1006, Fig. 7B (overlay of both arches on image acquired by mobile phone); EX1005, Fig. 7B; EX1002, ¶¶96-97.

Accordingly, as recited in step 1(a'), Salah and Carrier teach “activation of an image acquisition apparatus” (a mobile phone) “so as to acquire an image, called ‘analysis image,’ of said arch” (an image displayed on the screen during alignment, which is analyzed as addressed for subsequent limitations). EX1002, ¶98.

3. Step 1(b')

b') analysis of the analysis image by means of a deep learning device trained by means of a learning base;

Salah, Carrier, and Maninis teach step 1(b'). Salah teaches that its image acquisition apparatus may guide a user in acquiring images, including determining whether the image is acceptable as well as aiding in positioning via visual or audio feedback indicating how to position the acquisition apparatus. EX1008, 27:21-28:2. Salah further teaches that acquired images may be analyzed, detecting contours of teeth. EX1008, 34:5-14. A contour map of detected tooth edges is shown in Fig. 6b, reproduced below.

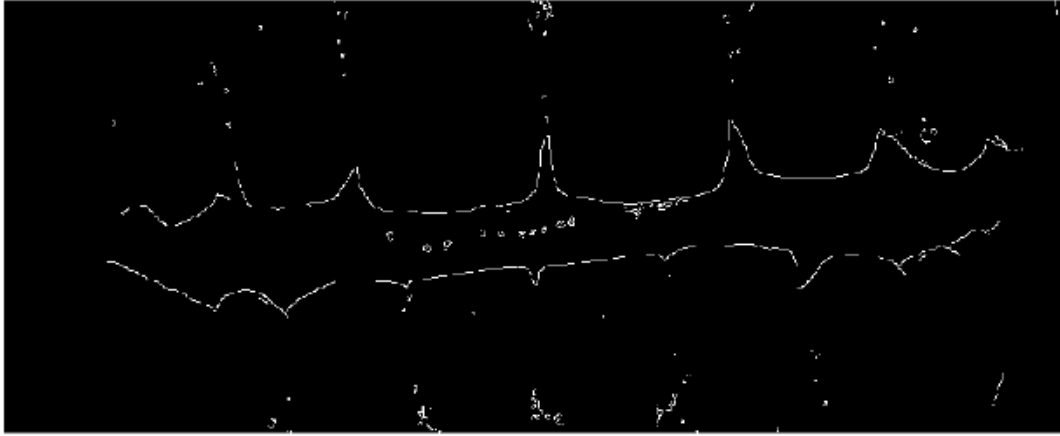


Fig. 6b

EX1008, 34:5-14, Fig. 6b; *see id.*, 34:5-36:14 (contour detection may employ deep learning); EX1002, ¶¶99-101.

Further, Carrier teaches that contour detection and object detection may be used to assess image alignment and guide users to acquire images. This includes analyzing images being acquired to ensure all teeth are visible and to detect the positions of tooth edges relative to an overlay. *See* EX1006, ¶¶63, 81; EX1005, ¶¶108, 128. Carrier’s techniques thus involve analyzing an acquired image. EX1002, ¶¶102-03.

Although Carrier does not expressly disclose that its image analysis techniques employ “a deep learning device trained by means of a learning base,” Carrier invites use of “any appropriate technique (e.g., edge detection image processing, etc.),” and Salah expressly identifies deep learning as an option for

detecting tooth contours. EX1006, ¶63; EX1005, ¶108; EX1008, 34:5-36:14; EX1002, ¶104.

Applicable deep learning techniques for contour and object detection are taught by Maninis. Maninis teaches methods of detecting edges (or “contours”) in images using a deep learning neural network to perform image analysis, using a process referred to as “Convolutional Oriented Boundaries (COB).” EX1007, Abstract. Maninis’s “contour detection” produces “multiscale oriented contours and region hierarchies” using “Convolutional Neural Networks (CNNs).” *Id.* Maninis describes improvements on “recent approaches to contour detection using deep learning,” presenting a “CNN architecture to produce contour detection at different scales, ... all in a single CNN forward pass at the image level.” EX1007, 3-5 (“contour probability map” used to detect region borders based on probability strength), Figs. 3-5 (showing results); EX1002, ¶105; *see also infra* §IV.D.3.

Maninis’s “deep learning architecture,” illustrated below, shows generation of contours using multiple feature map layers in a CNN to detect image contents at different scales (fine and coarse detections) as well as different orientations.

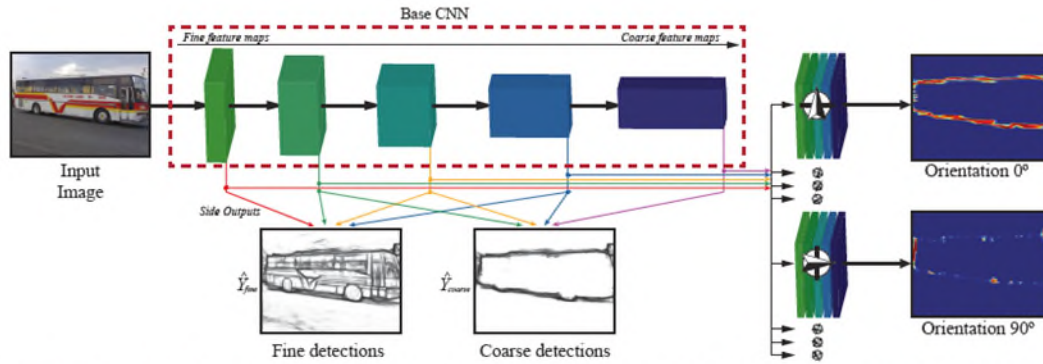


Fig. 2. Our deep learning architecture (best viewed in color). The connections show the different stages that are used to generate the multiscale contours. Orientations further require additional convolutional layers in multiple stages of the network.

EX1007, Fig. 2; EX1002, ¶106. As discussed in more detail for step 1(c'), Maninis's object and contour detection enables both detection of tooth edges and recognition of individual teeth as taught by Carrier. EX1002, ¶106; *infra* §IV.B.4.

Maninis's use of a computing device comprising a trained deep-learning convolutional neural network to analyze images matches the '409 patent's preferred embodiment. *E.g.*, EX1001, 1:24-25; EX1007, 3-5, 9; *see supra* §III.C; EX1002, ¶107. *Compare* EX1001, 16:15-42 (listing "Microsoft ResNet" and "VGG Net" as non-limiting embodiments), *with* EX1007, 2 ("[W]e present results both using VGGNet and ResNet....").

Maninis's deep learning device is trained on a "training dataset" (i.e., a "learning base" as recited in the '409 patent). EX1007, 3; EX1002, ¶108. In one example, Maninis uses a training dataset 10,103 images with "localized pixel-wise semantic annotations," meaning pixel-by-pixel indications of each object's location in each image. EX1007, 5-6; EX1002, ¶108. For training and optimization purposes, Maninis divides these images into "train, validation, and test sets" to

train and validate the deep learning device, substantially improving over prior classification techniques. *See* EX1007, 5-6, Table 3, Fig. 6; EX1002, ¶108.

Thus, Salah, Carrier, and Maninis teach “analysis of the analysis image” (to determine values for tooth presence and edges, as taught by Carrier) “by means of a deep learning device trained by means of a learning base” (a device comprising a deep learning neural network trained with a learning base, as taught by Salah and Maninis). EX1002, ¶109.

4. Step 1(c')

c') determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute;

Salah, Carrier, and Maninis teach step 1(c'). As noted above, an image attribute includes an attribute related to any of the categories listed in claim 13. *See supra* §III.A.

Salah's acquisition apparatus can provide feedback regarding whether an acquired dentition image is acceptable and indicate how the apparatus needs to be moved for correct positioning. EX1008, 27:21-28:2. While Salah does not expressly teach that this involves determination of a value for an image attribute, a POSA would have recognized that determining image attribute values would be a natural way to provide this functionality. EX1002, ¶¶110-11. Indeed, generating values for the image's relevant attributes would be a routine step to include in

software that determines whether an image is “acceptable” or properly positioned. *Id.* For example, as discussed below, at least two types of image analysis taught by Carrier involve determining such image attribute values. *Id.*

In one example of an image attribute, Carrier teaches analyzing an image to detect tooth edges, which are used to determine whether the teeth are in alignment by assessing a distance between the edges of each tooth and an overlay. EX1006, ¶63; EX1005, ¶108. EX1006, ¶63; EX1005, ¶108. For example, Carrier teaches that the distance may be measured to determine whether “the edges of the teeth are within” a threshold distance of the overlay. EX1006, ¶63; EX1005, ¶108; *cf.* EX1001, 26:16-24 (image attribute value may be set of values for multiple teeth). Alternatively, the measure may be an average of deviation from the overlay. EX1006, ¶68; EX1005, ¶114; EX1002, ¶112; *cf.* EX1001, 31:28-37 (image attribute value may be “mean separation” of teeth from aligner).

The set of distances from tooth edges to the overlay (or their average) constitutes a value for an image attribute because it relates to position or orientation of the acquisition apparatus (with larger distances indicating larger misalignment), as well as relating to the content of the analysis image (as distance of teeth from overlay relates to the representation of teeth or arches). Thus, each of Carrier’s distances is a value for an image attribute. EX1002, ¶113; *supra* §III.A.

Further, these distances are a function of step 1(b')'s analysis because Carrier's tooth edge distances are determined as a function of the location of edges, as derived from the deep learning contour detection taught by Salah and Maninis. EX1002, ¶114.

In another example of an image attribute, Carrier also discloses that the apparatus may use "image recognition algorithms to ensure that all required teeth are visible." EX1006, ¶81; EX1005, ¶128. A POSA would have understood that to ensure all required teeth are visible, each required tooth should be detected by the image recognition algorithm. EX1002, ¶115.

Both Carrier's distances and detections of required teeth are attributes of the image as a whole. Regarding distances, a set of all tooth distances or an average of those distances is an attribute of the image as a whole because each is a measure of overall image alignment, not merely a property of a single part of the image (e.g., one tooth) taken in isolation. Similarly, values representing whether the required teeth are present or absent would be image attributes because they represent a property of whether the required teeth are present in the image as a whole. EX1002, ¶121; *see* EX1001, 26:16-24, 31:28-37.

Maninis's deep learning techniques provide such an algorithm, because Maninis teaches how to use its contour detection algorithm to detect the presence of objects in an image. *See* EX1007, Abstract, 1-2, 11-12. Use of object detection,

as taught by Maninis, in the image acquisition apparatus of Salah and Carrier to detect visible teeth would be an image attribute related to the content of the analysis image—specifically, it would relate to which teeth are visible. EX1002, ¶116.

Maninis’s deep learning algorithm uses the contours it generates to define boundaries around regions and to identify the objects in each region of the image. See EX1007, Abstract, 11-12. Maninis describes such image recognition tasks as involving “image classification, object detection, and semantic segmentation” where detected contours are assigned to categories matching the “ground-truth” (i.e., identification what actual objects are where in an image). EX1007, 1-2. Objects are identified in regions by “label[ing] each of the detected pixels with the corresponding semantic class.” EX1007, 11-12; *see id.*, Tables 4-12 (identifying specific object types); EX1002, ¶117.

In view of Maninis’s teaching that its deep-learning-based contour detection techniques identify and classify objects in an image, a POSA would have been motivated to use those techniques to implement Carrier’s “image recognition algorithms” to detect specific teeth bounded by the detected contours. Adjusting Maninis’s deep learning device to perform the specific task of identifying teeth would have been straightforward and well within the skill of the POSA. EX1002, ¶118. These alterations would include providing an appropriately-annotated

training set and setting output layers to provide the required object identifications corresponding to the labels used to identify individual teeth, so that the identified teeth in an image could be compared to a set of required teeth. *Id.*

For example, a POSA would have been motivated to identify teeth at least by type, such as incisor, canine, pre-molar, or molar, as Salah and Carrier both show that these different types of teeth have different shapes. EX1002, ¶119; *see, e.g.*, EX1008, Fig. 4d (segmented arch model with different shapes for different tooth types); EX1006, Fig. 6B (different tooth shapes for both teeth and outlines); EX1005, Fig. 6B. To identify individual teeth, a POSA would have been further motivated to use the long-established system of tooth numbering to annotate each tooth in a training dataset used to train a neural network like Maninis’s to identify teeth by tooth number. EX1002, ¶119; *cf.* EX1009, ¶¶7-8, 12 (discussing known methods of categorizing teeth by type and/or number); EX1008, 54:23-26 (varying acquisition conditions based on tooth type, such as “incisor” and “molar”); EX1007, 5 (dataset annotated with 459 different semantic categories).

Because the deep learning techniques of Maninis support detection of objects such as the “required teeth” taught by Carrier, the values indicating whether the required teeth are present or not would likewise be determined as a function of the deep learning image analysis taught by Salah and Maninis. EX1002, ¶120.

Accordingly, Salah, Carrier, and Maninis teach step 1(c'). EX1002, ¶122.

5. Step 1(d')

d') comparison of said image attribute value with a setpoint;

Salah, Carrier, and Maninis teach step 1(d').

Salah's apparatus assesses whether an image is acceptable and properly positioned. EX1008, 27:21-28:2. While Salah does not expressly disclose that its evaluation involves comparison to a setpoint, a POSA would have considered such a comparison to be a routine option; for example, Carrier teaches the use of such comparisons of values to setpoints for determining whether an image is acceptable, as discussed below. EX1002, ¶¶123-24.

Regarding distance between tooth edge and overlay, Carrier teaches that these distances may be compared to a threshold, such as determining "whether the edges of the teeth are within about 1 mm, about 0.5 mm, about 0.1 mm, about 0.05mm, etc." of the overlay. EX1006, ¶63; EX1005, ¶108. If a sufficient percentage of teeth are closer than this threshold, this indicates that the apparatus is properly positioned. EX1006, ¶63; EX1005, ¶108; *see also* EX1006, ¶68 (threshold may use average distance), EX1005, ¶114. Thus, Carrier teaches that the "image attribute value" (distance from teeth to overlay) is compared with a setpoint (threshold distance for teeth to indicate alignment). EX1002, ¶125.

Regarding determining whether the image contains required teeth, Carrier also teaches or suggests comparison to a setpoint. Carrier teaches using image recognition to ensure “all required teeth are visible.” EX1006, ¶¶81; EX1005, ¶128. A POSA would naturally have applied Carrier’s teaching by determining whether every required tooth was visible or not by comparing a value (the required teeth’s visibility or non-visibility) to a setpoint (they must be visible). EX1002, ¶¶126-27.

Accordingly, Salah, Carrier, and Maninis teach step 1(d’).

6. Step 1(e’)

e’) sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof, 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

Salah, Carrier, and Maninis teach step 1(e’). Salah teaches using “written and/or voice prompts” to guide acquisition, such as by informing a user whether an image is acceptable. EX1008, 27:21-28:2. Salah also teaches guiding the user to acquire a new analysis image; for example, users can be told to “take another photo” when a first photo is unacceptable, and the device can “assist in positioning, for example by visual prompts (e.g. by displaying arrows), and/or audio messages

(such as succession of beeps, the frequency of which increases as device's positioning improves), and/or written and/or vocal prompts ('higher,' 'lower,' etc.).” *Id.* Thus, Salah teaches sending information messages as recited in step 1(e’); EX1002, ¶¶128-29.

Like Salah, Carrier teaches sending information messages regarding image alignment, with Carrier’s messages specifically checking whether the setpoint is respected. For example, Carrier teaches that the device may “indicate (e.g., by visual and/or audible and/or tactile) image that the teeth are aligned in the frame with the overlay” based on whether a sufficient number of teeth are sufficiently aligned to bring their edges within a threshold distance of the overlay. EX1006, ¶63; EX1005, ¶108; *see also* EX1006, ¶68, EX1005, ¶114. Carrier shows an indicator in Fig. 3B, explaining that “[i]f the overlay nearly matches the teeth, the indicator can be triggered, which indicates the moment to capture the image.” EX1006, ¶67; EX1005, ¶112. In Fig. 3B, the overlay changes color upon alignment, (changing from dark to light gray):

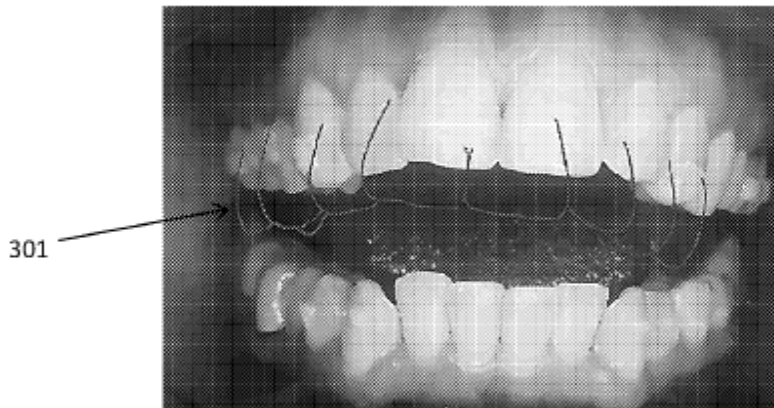


FIG. 3A

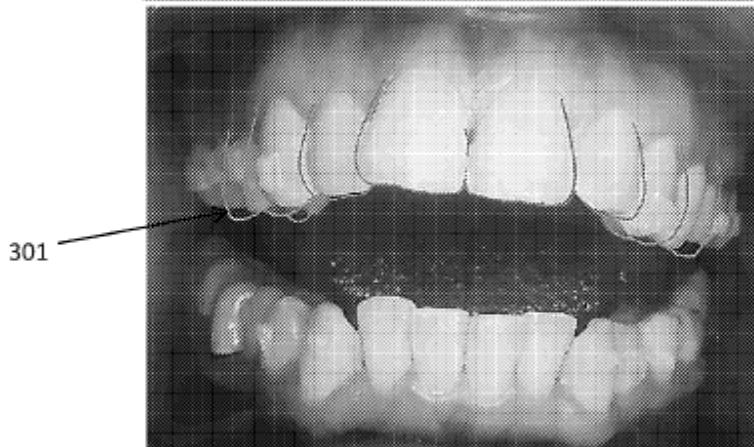


FIG. 3B

EX1006, Figs. 3A-B; *see also* EX1005, Figs. 3A-B; EX1002, ¶¶130-31.

Carrier further explains that the overlay may be colored red to indicate misalignment and green to indicate alignment. EX1006, ¶69; EX1005, ¶115. In another example, Carrier teaches that “one or more arrows may be shown on the screen to indicate that the user should move the mobile phone in a particular direction (or by a particular amount) to align the patient’s teeth with the overlay.” EX1006, ¶70; EX1005, ¶116. This resembles Salah’s use of arrows to indicate required position adjustments. EX1002, ¶132.

These messages also guide the user to acquire a new analysis image if the setpoint is not respected. The audible or visible indicators (such as color changes) guide a user because they indicate whether the alignment is correct. And the arrow indicators guide a user directly, as they show how to reposition the image acquisition apparatus. Thus, Salah and Carrier teach analogous indicators that guide a user to acquire a new image when a given image is not acceptable. EX1002, ¶133.

Likewise, Carrier's determination of whether all required teeth are visible would involve an information message based on a comparison that guides a user to acquire a new image when the original did not respect the setpoint. Carrier teaches using image recognition to "ensure that all required teeth are visible." EX1006, ¶81; EX1005, ¶128. Where not all required teeth are visible, this would render an image unacceptable, so a POSA would have been motivated to send an information message to this effect, such as the message taught by Salah that the user "must redo a photo." EX1008, 27:21-28:2. This message would relate to the comparison discussed for step 1(d') and it would expressly guide the user to acquire a new image by telling the user to redo imaging. EX1002, ¶134.

Finally, the messages taught by Salah and Carrier relate to the list of properties recited in step 1(e'). For example, the arrows and other audio or visual messages relate to "the position of the acquisition apparatus in relation to said

arch” because they indicate whether the acquisition apparatus (cellphone) is properly aligned with the patient’s dental arch or whether it needs to be adjusted, including optionally indicating the direction or amount of adjustment. EX1002, ¶135. The messages also relate to image quality and opening of the mouth. Salah expressly teaches messages regarding whether the image is “acceptable,” which is an assessment of quality. EX1008, 27:21-28:2; EX1002, ¶136. Where acceptability was based on the presence or absence of required teeth, as taught by Carrier, this would relate to quality (lack of visible teeth indicating poor quality), positioning (as mis-positioning can result in required teeth not being in frame), and opening of the mouth (as a closed mouth would obscure teeth). EX1002, ¶137; *see, e.g.*, EX1006, Figs. 9-10 (showing teeth no longer visible when mouth is closed); EX1005, Figs. 9-10.

Thus, Salah, Carrier and Maninis disclose sending an information message having each of the properties recited in of step 1(e’).

Accordingly, claim 1 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶¶138-39.

C. Claim 2

Claim 2 depends from claim 1 and further recites that “in the step b’), at least one analysis tooth zone is identified that represents, at least partially, a tooth on said analysis image, and at least one value of a tooth attribute is determined for

said at least one analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute value.” Salah, Carrier, and Maninis teach these elements. EX1002, ¶140.

Maninis teaches detecting contours to divide an image into regions, which can then be used for object detection. *See* EX1007, Abstract, 1-2, 11-12; *supra* §IV.B.3. In the context of Salah and Carrier, these contours would include the edges of teeth, and regions determined by these tooth edges would be regions determined to each contain a tooth. EX1002, ¶141. Each such region would thus be an identified “analysis tooth zone” that “represents, at least partially, a tooth on said analysis image.” *Id.*

Regarding determining tooth attributes for the zone, as discussed for claim 1, a POSA would have identified the teeth in the zone (e.g., by type or number) to accomplish Carrier’s teaching to use “image recognition algorithms to ensure that all required teeth are visible.” EX1006, ¶81; EX1005, ¶128. In combination with Salah and Maninis, this would involve detecting which tooth (by type/number) is present in each zone (i.e., each contour-defined region), because Maninis teaches that its regions may be used for object detection. *See* EX1007, Abstract, 1-2, 11-12. This identification would involve determination of a tooth attribute value for the zone. EX1002, ¶¶142-43; *supra* §III.B (scope of “tooth attribute”).

Further, using Maninis's edge detection to detect tooth edges meets another example of a "tooth attribute" as described in the '409 patent: "a tooth shape parameter." EX1001, 8:9-21. The edges of a tooth constitute a specific set of parameters describing the shape of the tooth, as the edges constitute the tooth's outline and are thus a literal description of the shape of the tooth in the image. EX1002, ¶144.

Finally, Salah, Carrier, and Maninis teach that "the value for the image attribute is determined as a function of said tooth attribute value." As discussed for claim 1, Carrier generates a value for distance based on the edges of each tooth, and also motivates generating values representing whether required teeth are present based on recognition of required teeth visible in the image. *See supra* §IV.B.4. Both of these values are generated based on a determined tooth attribute as recited in claim 2 (tooth edges or identity). EX1002, ¶145.

Accordingly, claim 2 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶146.

D. Claim 3

Claim 3 depends from claim 2 and further recites that step (b') comprises four additional steps. Salah, Carrier, and Maninis teach each of these steps, as discussed below. EX1002, ¶147.

1. Step 3(1)

1) creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image 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;

Regarding the number of images in a learning base, a POSA would have understood that a large number of images, often far more than 1000, would be typically used to train a deep learning device. EX1002, ¶144-29. For example, Maninis discloses training its neural network using over 10,000 images. *See* EX1007, 1, 5. Thus, a POSA would have found it natural to train a deep learning device such as a neural network with more than 1000 images. EX1002, ¶149.

A POSA would also have used historical “images of dental arches” when adapting Maninis’s contour and object detection to the dental imaging apparatus of Salah and Carrier. Maninis discloses using a training set with 459 semantic categories, corresponding to the objects depicted in the images. EX1007, 5; *see id.*, 11-12 (specific classification of objects); EX1002, ¶150. Training such a neural network teaches it to detect the categories of objects it has seen during training. EX1002, ¶150. Thus, to train a deep-learning-based neural network to recognize teeth and their edges, a POSA would have used images of patients’ teeth in the training set. *Id.* To obtain optimal results, a POSA would have used images

representative of the types of images likely to be seen during ordinary use of the imaging device, such as real historical photographs taken of patient's teeth with the device in which the trained network will be deployed (e.g., a cellphone). *Id.*

Finally, a POSA would have included zones representing teeth to which tooth attributes were assigned. Maninis teaches a learning base should include pixel-wise annotations indicating the regions for each object and labeling the types of objects (i.e., objects' attributes) that should be identified in that region. EX1007, 5-6; EX1002, ¶151. Specifically, Maninis's "pixel-wise semantic annotations" indicate which pixels correspond to each object in a given image, with objects' edges indicated by changes from one annotation to another. EX1002, ¶¶151-52. In the context of Salah and Carrier, a POSA would have applied annotations to historical dental images to indicate each object's attributes, including assigning tooth attributes to each tooth in each image. *Id.*

Thus, Salah, Carrier, and Maninis teach step 3(1).

2. Step 3(2)

2) training of at least one deep learning device by means of the learning base;

Maninis's deep-learning-based neural network is trained using a training dataset, which is a "learning base" in the context of the '409 patent. *See, e.g.*, EX1007, 3-4, 7-8; EX1002, ¶¶153-54. In the context of Salah and Carrier, a

learning base containing annotated historical images of dental arches would be used to train a neural network like Maninis's, thereby training "at least one deep learning device." EX1002, ¶155.

3. Step 3(3)

3) submission of the analysis image to the deep learning device for it to determine 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,

As discussed for claim 1, it would have been obvious to submit images of patients' dental arches, as captured by the device of Salah and Carrier, to be analyzed by a deep learning neural network as taught by Maninis to detect regions containing teeth and to characterize their attributes. EX1002, ¶¶155-57. Maninis teaches the determination of probabilities related to these functions. *Id.*

As Dr. Foroosh explains, Maninis's contour detection procedure generates a "contour detection probability map" that is used to determine "a hierarchical boundary map" demarcating its regions. EX1007, 3-5; EX1002, ¶157. In essence, Maninis describes determining a probability that a contour is located along a given set of pixels. EX1002, ¶157; *see* EX1007, 3 (providing equations to generate probability). Maninis uses these probability maps to derive the contours most likely to be present in the image (EX1007, 4-5), with different probabilities treated

as different “strengths” of boundaries. See EX1007, 4-5, Figs. 3-5 (strengths depicted by line thickness). These “strengths” are used to determine a “hierarchy of regions” dividing the different objects in an image. *Id.*; EX1002, ¶157.

In the context of Salah and Carrier, the probabilities relate to the presence of a zone at least partially representing a tooth because the edges corresponding to these probabilities define an analysis tooth zone that has been determined to contain a particular tooth. The attribute value is also “represented on said analysis tooth zone” because the attribute values discussed above (tooth identification and shape) are represented both visually as a property of the image as well as being assigned by Maninis’s image recognition process. EX1002, ¶¶158-59.

4. Step 3(4)

4) 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.

Maninis detects regions based on probabilities, as discussed above. See *supra* §IV.D.3; EX1007, 3-5. Thus, when a device based on the teachings of Salah, Carrier, and Maninis detects a tooth within such a region, that tooth’s presence and the corresponding tooth attributes would be detected as a function of that probability. EX1002, ¶¶160-61; see EX1006, ¶81; EX1005, ¶128.

Accordingly, claim 3 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶162.

E. Claim 4

Claim 4 depends from claim 3 and further recites “said tooth attribute is chosen from a tooth number, a tooth type, a tooth shape parameter, a tooth appearance parameter, a parameter relating to the state of the tooth, an age for the patient, or a combination of these attributes” As discussed above, Salah, Carrier, and Maninis teach or suggest using tooth attributes such as tooth type or number (to identify the required teeth taught by Carrier) or a tooth shape parameter (such as the tooth outlines taught by Salah and Carrier, detected as taught by Maninis). EX1002, ¶¶163-64; *see* EX1008, 34:11-14; EX1006, ¶¶63, 81; EX1005, ¶¶108, 128; EX1009, ¶¶7-8, 12; *supra* §§IV.C-D.

Accordingly, claim 4 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶165.

F. Claim 5

Claim 5 depends from claim 3 and further recites that “in the step 1) or 1’), a learning base is created comprising more than 10,000 historical images.” Claim 5 refers to a step “1’),” yet none of claims 1-3 includes a step “1’).” However, claim 3 includes a “step 1),” corresponding to one of claim 5’s alternatives.

Maninis discloses using more than 10,000 images in its training dataset.

EX1007, 1, 5-6. Using a similar number in the context of Salah and Carrier would be no more than a routine and predictable application of Maninis's teaching to use a sufficient number of training images (e.g., more than 10,000) to train the device of Salah and Carrier to detect tooth contours in images of patient's mouths.

EX1002, ¶¶166-67.

Accordingly, claim 5 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶168.

G. Claim 6

Claim 6 depends from claim 1 and further recites that “in the step b’), all said analysis tooth zones are identified, and at least one tooth attribute value is determined for each analysis tooth zone, and, in the step c’), the value for the image attribute is determined as a function of said tooth attribute values.” Although claim 6 states that “all said analysis tooth zones are identified,” neither claim 1 nor any prior language in claim 6 refers to an analysis tooth zone or indicates what set of analysis tooth zones is “all” being identified. Nonetheless, claim 2 refers to “at least one analysis tooth zone” being identified that “represents, at least partially, a tooth on said analysis image.” To the extent claim 6 is interpreted to address tooth zones like those recited in claim 2 for all members of some set of teeth depicted in an image, Salah, Carrier, and Maninis teach these elements. EX1002, ¶169.

That is, Petitioner assumes *arguendo* that claim 6 is satisfied at least if the elements of claim 2 are met for “all” tooth zones. Carrier teaches detecting edges of each tooth and ensuring that “all required teeth are visible.” EX1006, ¶¶63, 81; EX1005, ¶¶108, 128. As discussed above for claim 2, its elements are met for a given tooth based on the detection of tooth edges defining a detected region and identification of the particular tooth within that region. *See supra* §IV.C. Applying the same to each tooth in an arch comprising multiple teeth would meet the elements of claim 2 for each tooth, with the image attribute (e.g., presence of all teeth or distance from overlay) being determined from the teeth’s values, thus meeting each limitation of claim 6. EX1002, ¶¶170-73.

H. Claim 7

Claim 7 depends from claim 1 and further recites that step b’) comprises three additional steps. Salah, Carrier, and Maninis teach each of these steps, as discussed below in more detail. EX1002, ¶174.

1. Step 7(1’)

1’) creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image comprising an attribute value for at least one image attribute, or “image attribute value”;

Step 7(1’) is the same as step 3(1), except the historical images must comprise an image attribute value rather than historical tooth zones and tooth

attributes. A POSA would have been motivated to create a learning base of more than 1000 historical images of dental arches to train a neural network. *Supra* §IV.D.1. These images would comprise attribute values for at least one image attribute. For example, Maninis teaches pixel-wise annotations of all objects in each image of its training database. *See* EX1007, 5-6. In the context of Salah and Carrier, these images would be images representative of the type acquired for Salah and Carrier's dental images. EX1002, ¶¶175-77. Thus, for example, each acquired historical image would have a set of required teeth, corresponding to Carrier's teaching to ensure all required teeth are visible for a given image. *Id.*; *see* EX1006, ¶81; EX1005, ¶128. Accordingly, such an image would include annotations identifying whether all of the teeth required for that image were visible in the image, which would relate at least to the content of the image. *See supra* §§III.B, IV.D.1; EX1002, ¶177.

2. Step 7(2')

2') training of at least one deep learning device by means of the learning base;

This limitation is identical to step 3(2), so it is taught by Salah, Carrier, and Maninis for the reasons discussed in §IV.D.2. EX1002, ¶¶178-79.

3. Step 7(3')

3') 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.

As discussed for step 3(3), Maninis determines probabilities that it uses to detect boundaries between regions and enable identification of objects visible in those regions. EX1007, 4-5, 11-12; *see supra* §IV.D.3. The probabilities therefore relate to the image attributes discussed above for claim 1, such as whether required teeth are visible or the distance between the teeth's edges and Carrier's overlay. EX1002, ¶¶180-81.

Accordingly, claim 7 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶182.

I. Claim 8

Claim 8 depends from claim 1 and further recites that “the learning base is composed according to” six steps. Salah, Carrier, and Maninis teach each of these steps. In particular, the steps of claim 8 describe the use of a technique disclosed by Salah for registering a 2D image to a segmented arch model to map individual teeth to different portions of the image. *See* EX1008, Fig. 1. Attributes from the tooth model are assigned to the corresponding parts of the 2D image, which is then added to a learning base. As discussed below, this corresponds to Maninis's

teaching to include annotations reflecting the “ground truth” (i.e., Salah’s tooth model attributes) in training images. EX1002, ¶183.

A POSA would have been motivated to use Salah’s technique to identify locations of teeth in images to be added to a training library because doing so would enable the training dataset to contain “carefully localized pixel-wise semantic annotations” by providing a simple mechanism to connect regions of pixels in each image to the “ground-truth” represented by Salah’s tooth models. EX1007, 2, 5-6. Using Salah’s image-model registration technique would have enabled a POSA to leverage information contained in existing patient dental models to label images for training by labeling photographs acquired in conditions similar to those contemplated by Salah and Carrier. *See, e.g.*, EX1008, 1:20-21, 31:4-7 (tooth images acquired by patients’ mobile phones); EX1002, ¶¶184-85.

1. Step 8(A)

A) at an “updated” instant, production of a model of a dental arch of a patient, or “updated reference model”, and segmentation of the updated reference model so as to produce, for each tooth, a “tooth model”, and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model;

Salah teaches producing a model of a patient’s dental arch, which is segmented to provide a tooth model for each tooth. Salah describes “creation of a three-dimensional digital reference model of at least part of an arch, preferably at

least one of the patient's arches, or an 'initial reference model' and, preferably, for each tooth, definition, using the initial reference model, of a three-dimensional digital reference model of said tooth, i.e. the 'tooth model.'" EX1008, 1:28-32. Salah illustrates the process of creating the model, segmenting it, identifying each individual tooth, and providing a model for each such tooth. The initial model (prior to segmentation) is shown in Fig. 4a (below). See EX1008, 22:12-20, Fig. 4a; EX1002, ¶¶186-87.



Fig. 4a

Salah teaches segmenting the model by determining "which parts of the initial reference model correspond to which teeth ('tooth models')." EX1008, 23:4-8. This constitutes a segmentation of the tooth model, as it is a dividing of the

model into parts corresponding to different teeth, and it also provides a tooth model for each tooth as recited in step 8(A). EX1002, ¶188. Salah depicts a collection of tooth models in a segmented model of an entire dental arch in Fig. 4d, reproduced below:

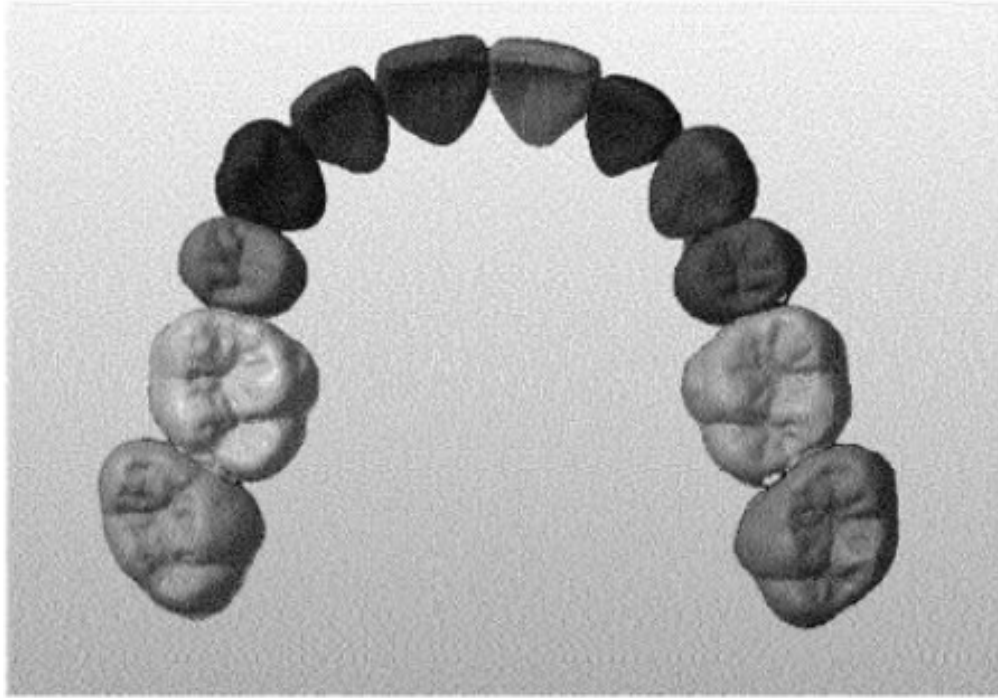


Fig. 4d

EX1008, Fig. 4d. This figure shows an individual tooth model for each tooth (shown by different shadings), confirming Salah's process produces a segmented reference model of a patient's dental arch with tooth models for each tooth.

EX1002, ¶188; see EX1001, Fig. 11D (same figure as Salah Fig. 4d), 7:9-16, 19:28-37 (describing this figure as a segmented arch model preferably generated as taught by Salah).

To the extent the term “updated” requires that the model be a second model, Salah teaches that both an initial and an updated model may be produced. *See, e.g.*, EX1008, 51:22-24 (“Comparison of the initial and updated reference models....”), 9:1-3 (describing generation of updated “reference model to be tested [that] can only be modified by moving one or more tooth models”), Fig. 1 (iterative process for updating and analyzing reference model); EX1002, ¶189.

Lastly, a POSA would have been motivated to assign a tooth attribute value to each tooth model. Salah’s generation of tooth models itself assigns tooth attribute values to the model. This is evident by inspection of Salah’s depiction of the models in Fig. 4d, showing different models (with different shape parameters) for different types of tooth. A POSA would have understood that when producing individual models of different teeth, the modeled teeth should be identified (e.g., by tooth number or type), since different teeth have different shapes, locations, and physical properties. *See, e.g.*, EX1009, ¶¶7-8, 12 (known to identify three-dimensional tooth models with attributes such as tooth type and number); EX1002, ¶190.

2. Step 8(B)

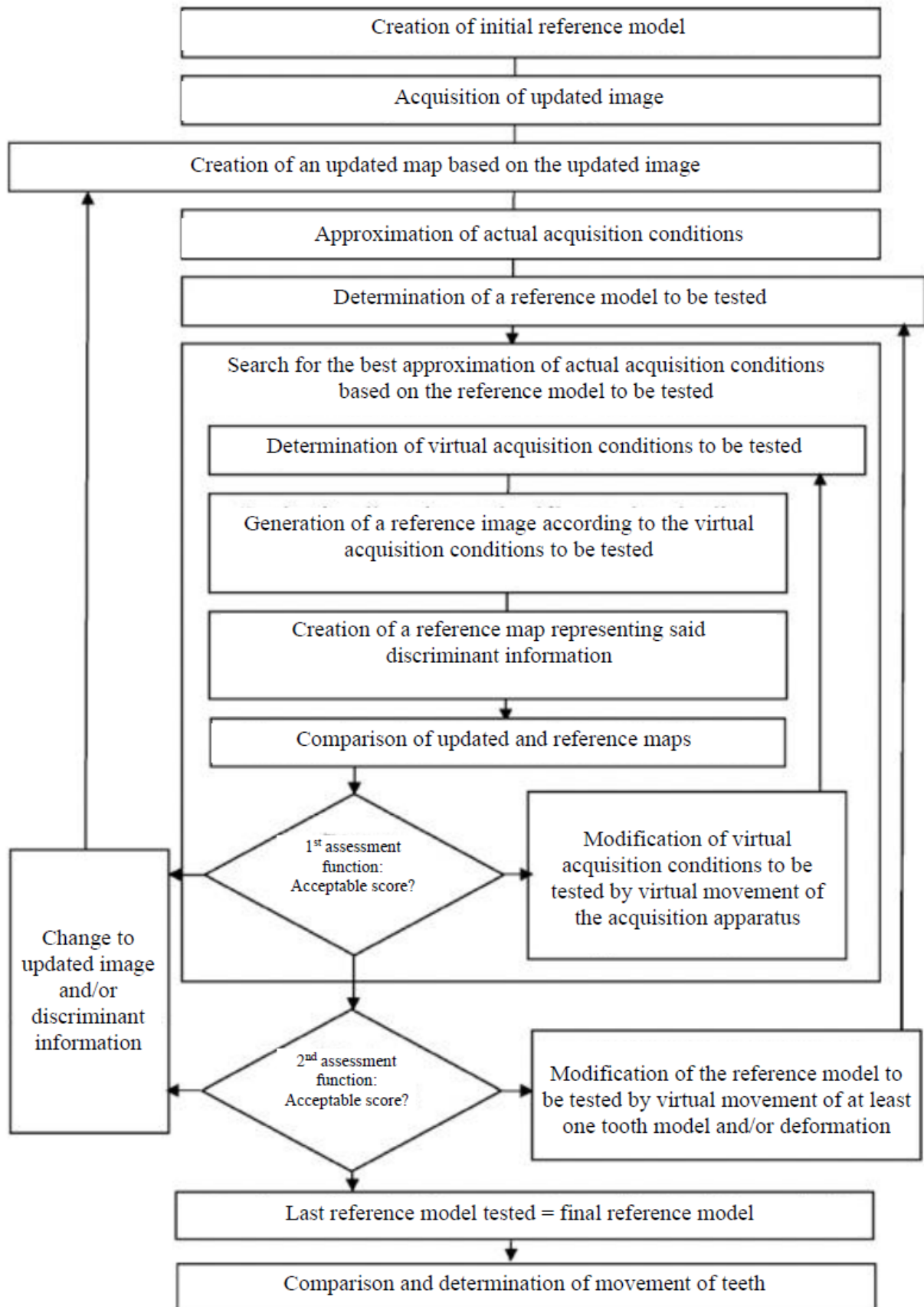
B) acquisition of at least one image of said arch, or “updated image”, in real acquisition conditions;

Salah teaches acquiring an updated arch image in real acquisition conditions. *See, e.g.*, EX1008, 2:1-2, 31:4-7, 32:1-2, 43:5-9, Fig. 1. The same applies to Carrier. *See, e.g.*, EX1006, ¶¶7, 59, 62-63, 94, Figs. 3A-B, 7B; EX1005, ¶¶8, 104, 107-08, 171, Figs. 3A-B, 7B. Acquiring updated images for inclusion in a learning base would be beneficial, as using real acquisition conditions would provide more realistic and effective training for a neural network as taught by Maninis. EX1002, ¶¶191-93; *see* EX1007, 5-6, Fig. 9-10, 16 (showing real images from training set).

3. Step 8(C)

C) for each updated image, search for virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called “reference image”, exhibiting a maximum match with the updated image in said virtual acquisition conditions, and acquisition of said reference image;

Salah describes matching an image to a reference model to obtain a reference image by finding “virtual acquisition conditions corresponding as closely as possible to real acquisition conditions.” EX1008, 2:28-3:30; *see id.*, 48:9-19 (finding virtual acquisition conditions for initial and updated reference models). This search is illustrated as part of the process in Fig. 1, below.



EX1008, Fig. 1. Thus, Salah describes a process meeting each limitation of step 8(C). EX1002, ¶¶194-95.

4. Steps 8(D)-(F)

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

Salah, Carrier, and Maninis teach these steps, which amount to using the correspondence between the updated image, reference image, and reference model to annotate the updated image with tooth attributes in each tooth zone, then add it to a learning base. EX1002, ¶196.

Because Salah’s dental model is composed of individual tooth models, identifying zones representing each tooth model in the reference image would be straightforward, as the locations of each tooth model used to generate the reference

image would be known. EX1002, ¶197. Salah likewise teaches detecting tooth contours in an updated image, as shown in Fig. 6b:

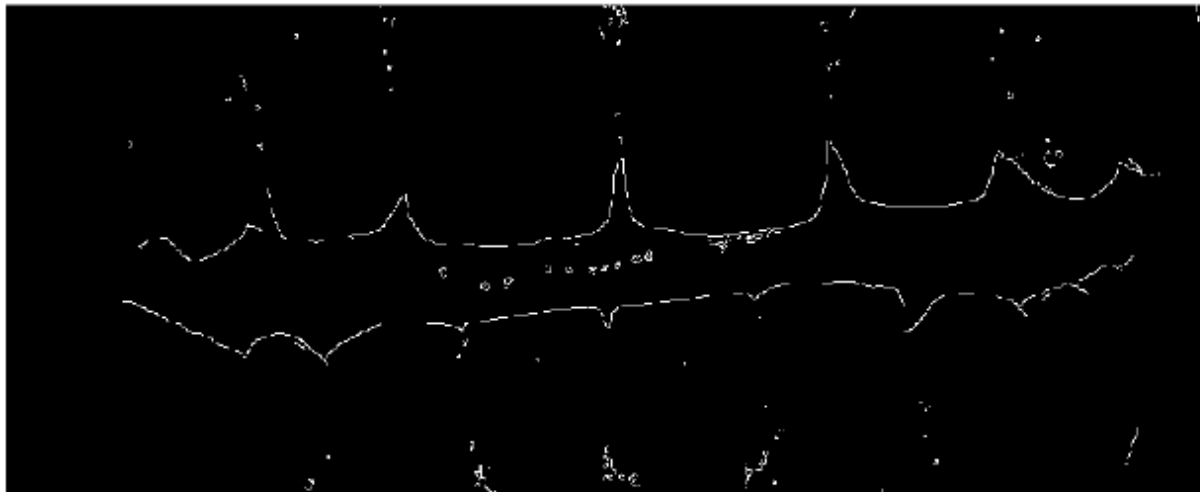


Fig. 6b

EX1008, Fig. 6b; *see id.*, 34:5-14; EX1002, ¶197. Using detected contours to form regions containing individual teeth would be useful because, as Maninis teaches, a training set should include pixel-wise annotations of which regions contain which objects, consistent with the “human ground-truth.” EX1007, 2, 5-6, 11-12; EX1002, ¶198.

As Dr. Foroosh explains, the creation of such annotations could be done in a straightforward fashion, either manually or with the assistance of automated tools. EX1002, ¶198. Salah teaches identifying “the area of the updated image containing the elements that were the subject of the initial reference model, i.e. the teeth and gums.” EX1008, 33:27-34:2. These identifications would allow accurate

annotations of images, as the correspondence between the reference and updated images would mean that the identifications would map from the reference model and image to the updated image, thus providing efficient annotations to the updated image. EX1002, ¶¶198-99 This process would merely be the application of Salah’s teaching to register arch models with images to Maninis’s teaching to provide annotations of all objects in an image to be included in a training dataset, according to known techniques for projecting annotations from a 3D model to a matching 2D image. *Id.*; *see, e.g.*, EX1013, 3689-90; EX1014, 3198, 3200-01. This would produce a predictable result: images would be added to a training dataset with annotations matching the ground-truth represented by Salah’s model. EX1002, ¶199.

Thus, Salah, Carrier, and Maninis teach steps 8(D)-(F).

Accordingly, claim 8 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶200.

J. Claim 9

Claim 9 depends from claim 1 and further recites that “the learning base is composed according to” six steps, each of which either resembles or is identical to the steps in claim 8. Salah, Carrier, and Maninis teach each of these steps, as discussed below. EX1002, ¶201.

1. Step 9(A')

Step 9(A') is identical to step 8(A), except “updated” in claim 8 is replaced throughout with “initial.” As step 9(B') refers to an “updated” instant, claim 9 appears to require that steps A' and B' be performed in order. Consistent with this ordering, Salah teaches that both initial and updated models may be generated, segmented, and compared with updated images. *See, e.g.*, EX1008, 1:28-2:5, 23:4-8, Fig. 4d. Indeed, Salah's teachings cited above for claim 8(A) are applicable both to “initial” and “updated” models. *See id.*; EX1002, ¶202. Thus, step 9(A') is taught by Salah, Carrier and Maninis for similar reasons to step 8(A). EX1002, ¶203; *supra* §IV.I.1.

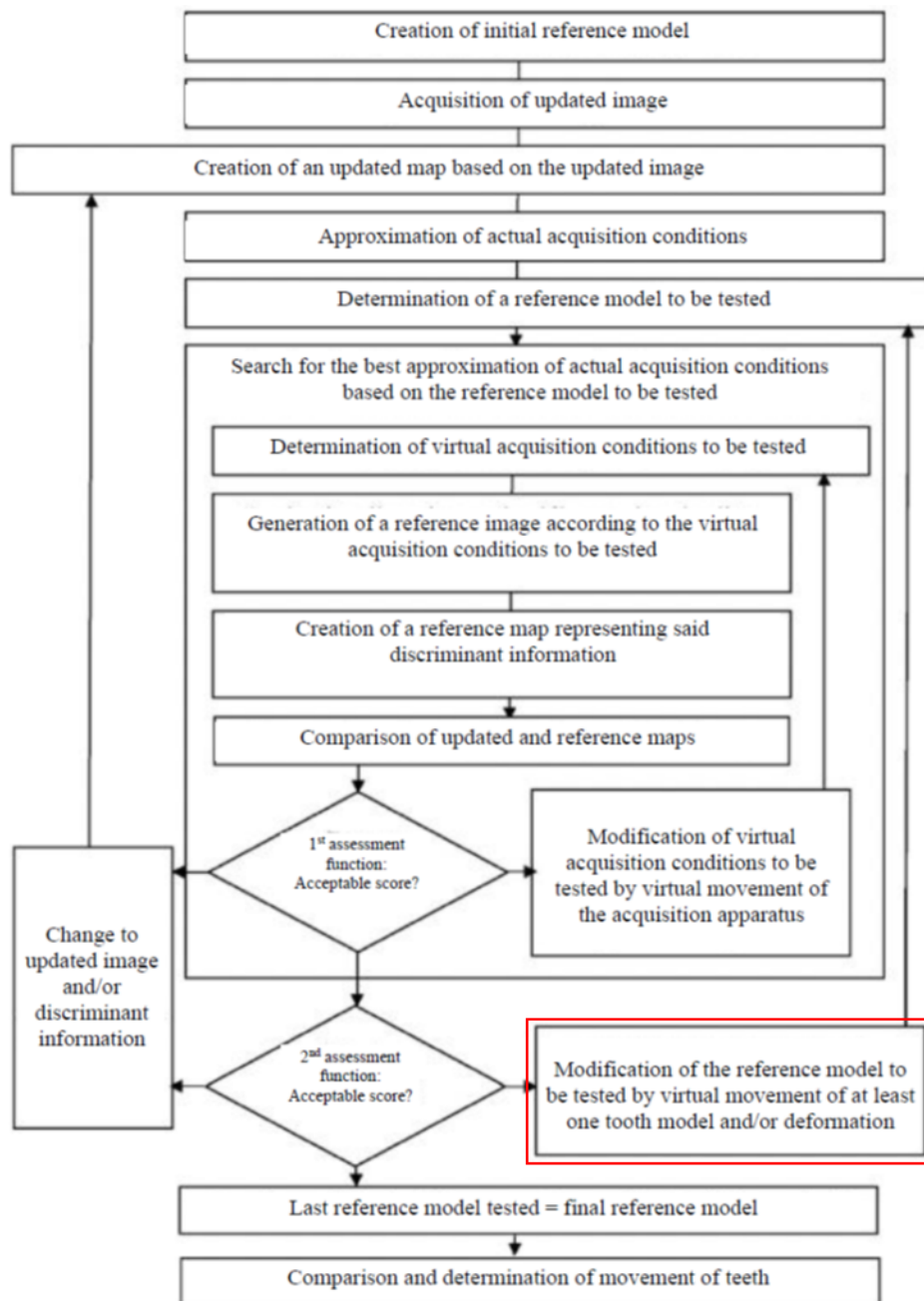
2. Step 9(B')

Step 9(B') is identical to step 8(B), except Step 9(B') begins “at an updated instant,” whereas in claim 8 the updated instant is recited in step 8(A). Salah expressly describes using an “initial” model in combination with an “updated” image. EX1008, 1:28-2:5. Thus, acquiring the “updated image” is similarly consistent with Salah's disclosure. Accordingly, Step 9(B') is taught by Salah, Carrier, and Maninis for similar reasons to step 8(B). EX1002, ¶¶204-05; *supra* §IV.I.2.

3. Step 9(C')

Step 9(C') recites “search, by deformation of the initial reference model, for an updated reference model and virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called ‘reference image,’ exhibiting a maximum match with the updated image in said virtual acquisition conditions.” Step 9(C') omits two limitations from step 8(C): “acquisition of said reference image” and searching “for each updated image.” The analysis for step 8(C) applies to step 9(C') notwithstanding these omissions. EX1002, ¶¶206-07. Step 9(C') otherwise differs from step 8(C) only by adding a limitation that searching be done “by deformation of the initial reference model, for an updated reference model.” This added limitation is taught by Salah. *Id.*

Salah teaches generating updated models by deformation of an initial model, such as by adjustment of individual modeled teeth. Salah describes that its sequence of repeated optimizations of a reference model, including the steps discussed above for claim 8, may further include iteratively deforming a model. EX1008, 49:13-24. This step is illustrated in Fig. 1 of Salah as being part of the matching process between an updated image and a reference model. This figure is shown below, with the deformation step marked by a red box:



EX1008, Fig. 1 (annotated). Salah explains that such deformations allows generation of “a final reference model that takes into account both tooth displacement and tooth deformation.” EX1008, 49:5-7. Salah thus teaches how to account for changes of teeth over time (such as position changes) by deforming its

model as part of the process discussed above for claim 8(C) to match an updated image. EX1002, ¶208.

Accordingly, Step 9(C') is taught by Salah, Carrier, and Maninis. *See supra* §IV.I.3; EX1002, ¶209.

4. Steps 9(D)-(F)

Steps 9(D)-(F) are respectively identical to steps 8(D)-(F). A POSA would have performed these steps in the context of the elements recited in claim 9 for the same reason as for the similar elements in claim 8. Accordingly, Salah, Carrier, and Maninis teach steps 9(D)-(F) for the reasons discussed for steps 8(D)-(F). EX1002, ¶210; *supra* §IV.I.4.

Accordingly, claim 9 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶211.

K. Claim 10

Claim 10 depends from claim 1 and further recites “in which, to create a historical image of the learning base, identifies one or more ‘historical’ tooth zones on an image, called then assigns, to each identified historical tooth zone, a value for at least one tooth attribute, and/or assigns to an image a value for at least one tooth attribute.” Claim 10 does not describe who or what “identifies” or “assigns” as recited, but for purposes of this petition, Align assumes it is the historical image’s creator. Second, it is not clear what is meant by the phrase “called then

assigns” in the first recited alternative; however, regardless of the scope of the first alternative, Salah, Carrier, and Maninis teach the second alternative, which is sufficient to find the claim unpatentable. *Intel Corp. v. Qualcomm, Inc.*, 21 F.4th 801, 813 (Fed. Cir. 2021).

Maninis teaches its images contain “localized pixel-wise semantic annotations for the entire image” (EX1007, 5-6), which in the context of Salah and Carrier would indicate which objects were in each region of the image, such as identifications of which specific teeth (or tooth types) are visible in each region of the image. *See supra* §IV.D. To produce such annotations, a POSA would have assigned a tooth attribute value (e.g., tooth types or numbers) to each tooth-containing region (tooth zone). Because each tooth zone is part of an image, this would be an assignment of a value for a tooth attribute to the image, thus satisfying at least claim 10’s second alternative. EX1002, ¶¶212-14.

Accordingly, claim 10 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶215.

L. Claim 11

Claim 11 depends from claim 1 and further recites that “the acquisition apparatus is a cellphone.” Salah and Carrier both teach that the acquisition apparatus is a cellphone. *E.g.*, EX1008, 12:7 (“the acquisition apparatus is a cell

phone”), 31:4-7; EX1006, ¶¶30, 62, Figs. 2A-2C, 5A-7C; EX1005, ¶¶66, 99, Figs. 2A-2C, 5A-7C; EX1002, ¶216-17.

Accordingly, claim 11 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶218.

M. Claim 12

Claim 12 depends from claim 1 and further recites that “the information message is sent by the acquisition apparatus.” Salah and Carrier both teach that the information message is sent by the acquisition apparatus. *E.g.*, EX1008, 27:21-28:2 (apparatus provides “written and/or vocal prompts”); EX1006, ¶¶63 (“visual and/or audible and/or tactile” feedback), 69-70 (apparatus changes overlay color, shows arrows); EX1005, ¶¶108, 114-15.

Accordingly, claim 12 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶¶219-21.

N. Claim 13

Claim 13 depends from claim 1 and further recites that the “image attribute relates to” any of several categories. Salah, Carrier, and Maninis teach these elements. EX1002, ¶222.

One category listed in claim 13 is that the image attribute may relate to “a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image.” As discussed above for claim 1 (*supra* §IV.B.4),

Carrier teaches determining, from detected tooth edges, the distance between teeth and an overlay, either as a set of values for each tooth, or an average of distance values, either of which is an image attribute value related to the apparatus's position or orientation. Indeed, Carrier uses this measurement to determine whether the apparatus is aligned. EX1006, ¶¶63, 68; EX1005, ¶¶108, 114; EX1002, ¶223.

Another category in claim 13 is an image attribute related to the content of the image, including representation of, among other options, “the arches” or “one or more teeth.” As discussed for claim 1 (*supra* §IV.B.4), Carrier teaches determining whether “all required teeth are visible,” which a POSA would have accomplished using an image attribute value representing whether the required teeth (such as an arch, or a subset of specified teeth) were all visible. EX1006, ¶81; EX1005, ¶128; EX1002, ¶224.

Accordingly, claim 13 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶225.

O. Claim 14

Claim 14 depends from claim 1 and further recites that “said analysis image is analyzed and checked in real time.” Salah, Carrier, and Maninis teach these elements. EX1002, ¶226.

Carrier discloses conducting its image analysis and checking (e.g., comparison to setpoints) in real time. *E.g.*, EX1006, ¶63 (“automatic detection of alignment in real-time with a patient’s teeth and an overlay”); EX1005, ¶108; *see also* EX1006, ¶54 (overlay on real-time image with immediate feedback); EX1005, ¶99. Thus, Carrier teaches that its image analysis and checking is performed in real time. EX1002, ¶227.

The combination of Salah and Maninis with Carrier to use a deep learning device for the analysis is consistent with Carrier’s teaching of real-time analysis. Like Carrier, Salah also provides real-time feedback on alignment. *See* EX1008, 27:21-28:2. Maninis teaches that its deep learning contour detection algorithms produce fast, efficient, and highly-accurate results, categorizing even challenging images in less than 800 ms. EX1007, 1-2; *see id.*, 9, Table 1. Thus, a POSA would have expected Maninis’s deep-learning-based neural network to maintain Carrier’s capacity for real-time image analysis. EX1002, ¶228.

Accordingly, claim 14 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶229.

P. Claim 15

Claim 15 depends from claim 1 and further recites that “the deep learning device is a neural network.” Maninis teaches that its deep learning image analysis for contour and object detection is performed using a convolutional neural network

(CNN). EX1007, Abstract; *see also id.*, 1-2, 13, Figs. 1-2. Accordingly, claim 15 would have been obvious over Salah, Carrier, and Maninis. EX1002, ¶230-32.

For the foregoing reasons, claims 1-15 would have been obvious over Salah, Carrier, and Maninis. ¶EX1002, ¶233.

V. DISCRETIONARY DENIAL IS NOT WARRANTED

This petition does not implicate the Board's discretion according to *Fintiv*. *Apple Inc., v. Fintiv, Inc.*, IPR2020-00019, Paper 11. *See generally* Memorandum on Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation (June 21, 2022) (*Fintiv Memo*). Patent Owner filed its district court complaint in the Northern District of California on November 18, 2022, asserting three patents including the '409 patent. At a case management conference on July 14, 2023, the court ordered the parties to proceed with discovery relating to two claims from the three asserted patents, the claims to be selected by September 25, 2023, and the court scheduled a summary judgement hearing on those claims for February 22, 2024. After the summary judgment hearing, the parties expect to proceed with discovery relating to remaining issues in the case.

The district court proceedings remain in early stages. For example, this petition was filed less than two months after the initial case management conference, barely one month after Patent Owner's service of infringement

contentions, and prior to the asserted claims being finalized. Expert reports have not been exchanged, no work on claim construction has been completed, and no Markman hearing has been scheduled. While the parties have agreed to be ready for trial by November 2024, no trial date has been scheduled. EX1027.

Nonetheless, the median time from filing to trial in the Northern District of California is 37.1 months. EX1020, 66. The expected trial date is therefore late December 2025, whereas the expected date of a final written decision here would be February or early March 2025. *See Fintiv* Memo, 8-9 (institution favored where median time-to-trial indicates earlier FWD date). Moreover, the petition presents compelling merits, as discussed in detail in §IV.

This petition also does not implicate §325(d). The petition presents new art and argument, as neither Carrier nor Maninis was disclosed or considered during examination, and although Salah was cited in an IDS and in the specification of the '409 patent, the version submitted was in French, with only an English abstract, not an English translation. *See* EX1004, 246-322, 840, 1009. Nor are the references relied on in the petition cumulative to the references applied during examination, as the discussion of claim 1 shows that the elements indicated as missing by the Examiner (i.e., steps 1(d') and 1(e'), *see* EX1004, 868-69) were taught in the Carrier and Salah references. Finally, the discussion in the ground above explains

in detail why the claims of the '409 patent were obvious over the prior art, confirming that their allowance constituted material error. *See supra* §IV.

VI. GROUNDS FOR STANDING

Petitioner certifies that, under 37 C.F.R. §42.104(a), the '409 patent is available for *inter partes* review, and Petitioner is not barred or estopped from requesting *inter partes* review of the '409 patent on the grounds identified.

VII. CONCLUSION

For the reasons set forth above, claims 1-15 of the '409 patent are unpatentable. Petitioners therefore request that a *inter partes* review of these claims be instituted.

Respectfully submitted,

Dated: August 31, 2023

/ Michael T. Rosato /
Michael T. Rosato, Lead Counsel
Reg. No. 52,182

VIII. MANDATORY NOTICES UNDER 37 C.F.R. §42.8

Real Party-in-Interest (37 C.F.R. §42.8(b)(1)): Align Technology, Inc. is the real party-in-interest.

Related Matters (37 C.F.R. §42.8(b)(2)): Patent Owner has asserted the '409 patent against Petitioner in the United States District Court in the Northern District of California (3-22-cv-07335).

Lead and Back-Up Counsel (37 C.F.R. §42.8(b)(3))

Lead Counsel: Michael T. Rosato (Reg. No. 52,182)

Back-Up Counsel: Matthew A. Argenti (Reg. No. 61,836); Patrick M. Medley (Reg. No. 74,735)

Service Information – 37 C.F.R. §42.8(b)(4). Petitioner hereby consents to electronic service.

Email: mrosato@wsgr.com; margenti@wsgr.com; pmedley@wsgr.com

Post: Wilson Sonsini Goodrich & Rosati, 701 Fifth Avenue, Suite 5100,
Seattle, WA 98104-7036

Tel.: 206-883-2529

Fax: 206-883-2699

IX. PAYMENT OF FEES UNDER 37 C.F.R. §§42.15(A) AND 42.103

The required fees are submitted herewith. If any additional fees are due at any time during this proceeding, the Office is authorized to charge such fees to Deposit Account No. 23-2415.

X. CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this Petition complies with the type-volume limitation of 37 C.F.R. §42.24(a). The word count application of the word processing program used to prepare this Petition indicates that the Petition contains 13,993 words, excluding the parts of the brief exempted by 37 C.F.R. §42.24(a).

Dated: August 31, 2023

/ Michael T. Rosato /
Michael T. Rosato, Lead Counsel
Reg. No. 52,182

XI. APPENDIX – LIST OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 10,755,409 to Salah et al.
1002	Declaration of Dr. Hassan Foroosh
1003	Curriculum Vitae of Hassan Foroosh, Ph.D.
1004	U.S. Patent Application No. 16/031,201 File History
1005	U.S. Patent Pub. No. 2021/0068923 to Carrier et al.
1006	Provisional Patent Application No. 62/417,985 to Carrier et al.
1007	Maninis et al., “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks,” <i>arXiv.org</i> (Apr. 28, 2017)
1008	Translation of PCT Pub. No. WO 2016066651 to Salah et al.
1009	U.S. Patent Pub. No. 2013/0282351 to Martin Tank
1010	U.S. Patent Pub. No 2020/0229901 (Application No. 16/827,594)
1011	U.S. Patent Pub. No. 2018/0125610 (Application No. 15/803,718)
1012	PCT Pub. No. WO 2016066651 to Salah et al.
1013	Xie et al., “Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer,” <i>2016 IEEE Conference on Computer Vision and Pattern Recognition</i> (2016)
1014	Chen et al., “Beat the MTurkers: Automatic Image Labeling from Weak 3D Supervision,” <i>2014 IEEE Conference on Computer Vision and Pattern Recognition</i> (2014)
1015	Krizhevsky et al., “ImageNet Classification with Deep Convolutional Neural Networks,” <i>Advances in Neural Information Processing Systems</i> 25 (2012)
1016	Patrick Cavanaugh, “Visual Cognition” <i>Vision Research</i> 51 (2011)

1017	Prasad et al., “Learning Class-specific Edges for Object Detection and Segmentation,” <i>Computer Vision, Graphics and Image Processing: 5th Indian Conference, ICVGIP 2006</i> (2006)
1018	Wang et al., “Simultaneous Image Classification and Annotation,” <i>CVPR 1</i> , No. 2 (2009)
1019	Tao et al., “Similarity Voting based Viewpoint Selection for Volumes,” <i>Computer Graphics Forum</i> 35, No. 2 (2016)
1020	United States District Courts — National Judicial Caseload Profile, available at https://www.uscourts.gov/sites/default/files/data_tables/fcms_na_distprofile0331.2023.pdf (March 2023)
1021	Maninis et al., “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks,” <i>arXiv.org</i> (Jan. 17, 2017)
1022	<i>arXiv.org Submission Help Page</i> , https://arxiv.org/help/submit (as archived May 3, 2017, at https://web.archive.org/web/20170503071322/https://arxiv.org/help/submit)
1023	<i>arXiv.org Article Replacement Help Page</i> , https://arxiv.org/help/replace (as archived July 18, 2017, at https://web.archive.org/web/20170718173822/https://arxiv.org/help/replace)
1024	<i>COB - Convolutional Oriented Boundaries</i> , http://people.ee.ethz.ch/~cvlsegmentation/cob/ (as archived May 3, 2017, at http://web.archive.org/web/20170503233424/http://people.ee.ethz.ch/~cvlsegmentation/cob/)
1025	<i>arXiv.org Author Search Page for Luc van Gool</i> , https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1 (as archived July 8, 2017, at https://web.archive.org/web/20170708002617/https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1)
1026	<i>arXiv.org Article Replacement Help Page</i> , https://arxiv.org/help/replace (as archived September 8, 2016, at https://web.archive.org/web/20160908205209/https://arxiv.org/help/replace)

1027	Joint Stipulation and Proposed Order (Dkt. No. 73), <i>Dental Monitoring SAS v. Align Technology, Inc.</i> , No. 3-22-cv-07335-WHA (N.D. Cal.)
------	--

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§42.6(e) and 42.105(a), this is to certify that I caused to be served a true and correct copy of the foregoing Petition for Inter Partes Review of U.S. Patent No. 10,755,409 (and accompanying Exhibits 1001-1027) by overnight courier (Federal Express or UPS), on this 31st day of August, 2023, on the Patent Owner at the correspondence address of the Patent Owner as follows:

Cooper Legal Group LLC
1388 Ridge Road, Unit 1
Hinckley, OH 44233

Dated: August 31, 2023

/ Michael T. Rosato /
Michael T. Rosato, Lead Counsel
Reg. No. 52,182