

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ALIGN TECHNOLOGY, INC.,
Petitioner,

v.

DENTAL MONITORING SAS,
Patent Owner.

Case No. IPR2024-00052
Patent No. 11,049,248

DECLARATION OF DR. HASSAN FOROOSH

ALIGN EXHIBIT - 1002

TABLE OF CONTENTS

I.	QUALIFICATIONS	1
II.	SCOPE OF WORK.....	6
III.	LEGAL STANDARDS	7
IV.	BACKGROUND IN THE ART	11
	A. Image Analysis and Computer Vision	11
	B. Edge Detection and Contour Detection.....	14
	C. Deep Learning	15
V.	THE '248 PATENT	17
VI.	PRIOR ART.....	20
	A. Kim	20
	B. Invisalign	21
	C. Salah	23
	D. Maninis	25
VII.	OVERVIEW OF THE LEVEL OF SKILL.....	29
VIII.	CLAIM CONSTRUCTION	30
	A. tooth attribute	31
	B. image attribute	33
	C. image extracted from a film	36
IX.	CLAIMS 1-2, 4, 7-8, AND 13-16 WERE OBVIOUS AS ILLUSTRATED BY KIM, INVISALIGN, SALAH, AND MANINIS	36
	A. The Combination of Kim, Invisalign, Salah, and Maninis	36

B.	Claim 1	47
1.	Claim 1 preamble	47
2.	Step 1(a'')	48
3.	Step 1(b'')	55
C.	Claim 2	67
D.	Claim 4	69
E.	Claim 7	71
1.	Step 7(1).....	71
2.	Step 7(2).....	75
3.	Step 7(3).....	76
4.	Step 7(4).....	79
F.	Claim 8	80
1.	Step 8(1')	80
2.	Step 8(2')	82
3.	Step 8(3')	83
G.	Claim 13	84
H.	Claim 14	85
I.	Claim 15	87
J.	Claim 16	90
X.	CONCLUDING STATEMENTS.....	93
XI.	APPENDIX – LISTING OF CLAIMS.....	94
XII.	APPENDIX – EXHIBITS	97

I, Hassan Foroosh, declare as follows:

I. QUALIFICATIONS

1. I have been retained by counsel on behalf of Align Technology, Inc. to provide opinions in this matter related to my experience and expertise. My opinions are based on my skills, knowledge, training, education, and experience in matters of this nature, and my examination of the materials utilized in preparing this report.

2. I am a Professor of Computer Science at the University of Central Florida. My primary research interests include image/video processing, computer vision, machine learning, and image-based modeling, including 3D modeling.

3. I received my Ph.D. in Computer Science from INRIA (French National Institute for Research on Informatics and Control), University of Nice Sophia Antipolis in 1996. I also earned a Master of Science in Computer Science from the University of Nice Sophia Antipolis in 1993, a Master of Science in Electrical Engineering from Kingston University, U.K., in 1990, and a Bachelor of Science in Electrical Engineering from Kingston University, U.K., in 1989.

4. Prior to research and teaching, I worked as an R&D Engineer at Innovative Engineering Solutions Ltd. in London from 1990 to 1992. Between 1993 and 1996, I served as a Research Assistant in the Scene Analysis and Symbolic Image Processing Group of INRIA, where I performed research in areas

such as 3D super-resolution, imaging processes for registration, and shape perception. From 1997 to 2000, I served as an Assistant Research Scientist at the University of Maryland, College Park, where I investigated topics such as image registration (2D and 3D), super-resolution of video data for unmanned aerial vehicles, super-resolution of Synthetic Aperture Radar (SAR) and Foliage Penetrating SAR (FOPEN-SAR), Automatic Target Detection and Tracking (ATD/R). During this time, I authored one of the most cited works on registration and alignment based on phase correlation: (Hassan Foroosh (Shekarforoush), Josiane Zerubia and Marc Berthod, Extension of Phase Correlation to Sub-pixel Registration, IEEE Transactions on Image Processing, vol. 11, Issue 3, pp. 188-200, 2002). Between 2000 to 2002, I served as Senior Research Scientist at the University of California, Berkeley where I also served as the Technical Coordinator and Principal Research Scientist of a Multi-University Research Initiative (MURI) funded by the Office of Naval Research. As part of the project, I conducted research on multi-sensor acquisition of 3D data for autonomous 3D modeling and visualization.

5. I joined the Department of Computer Science at the University of Central Florida (UCF) as an Assistant Professor in 2002. Since then I have served as an Associate Professor, Professor, and CAE Link Professor of Computer Science in the same department. I also served in various capacities in the

department, such as the Graduate Program Coordinator, chair of various leadership committees, and most recently in the Faculty Senate Strategic Planning Council for UCF. I have taught a number of courses directed to computer vision and 3D modeling, including: 3D Computer Vision, Image Processing, Special Topics in 3D Computer Vision, and Visual Simulation.

6. I have conducted extensive research in the field of 3D computer vision and computer modeling. Some example research projects that I have conducted include *Video Analysis Techniques for Terrain Modeling and Automated Texture Mapping* (2006-2007) and *Modeling and Visualization of Musculoskeletal Structures in Dental CT-Scan* (2006-2007). For the latter project, I developed algorithms for 3D registration, modeling and visualization of maxilla and mandible of patients for dental surgery using Computer Aided Tomography (CAT) scanning.

7. I have been the principal investigator or co-principal investigator of numerous research projects funded by the National Science Foundation (NSF), NASA, DoD, and industry, with total funding exceeding \$14M since joining UCF. As one example, I recently served as the principal investigator in two projects funded by the National Science Foundation (NSF) on 3D modeling and the role of 3D information in machine understanding of digitized world. One project was directed to the registration/alignment of dental CT scans of patients over a period

of time during the treatment of a patient, with the purpose of improving the quality of 3D models, e.g., corrections and noise removal in 3D CT scans.

8. I have authored or co-authored over 180 articles in peer reviewed journals or conference proceedings. Some example publications include M. Berthod, H. Shekarforoush (Foroosh), M. Werman and J. Zerubia, *Reconstruction of High Resolution 3D Visual Information*, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages. 654-657 (1994); C. Früh, J. Flynn, H. Foroosh, and A. Zakhor, *Fast 3D model generation for urban environments*, Workshop on the Convergence of Graphics, Vision, and Video (CGVV'01), Berkeley, March 2001; Jiangjian Xiao, Xiaochun Cao, and Hassan Foroosh, *3D Object Transfer Between Non-Overlapping Videos*, Proceedings of IEEE Virtual Reality Conference, pages 127-134, (March 25-29, 2006); M. Balci, M. Alnasser, and H. Foroosh, *Subpixel Alignment of MRI Data Under Cartesian and Log-Polar Sampling*, Proceedings of IAPR International Conference on Pattern Recognition (ICPR), vol.3, pages 607-610 (2006); M. Balci, M. Alnasser, and H. Foroosh, *Alignment of Maxillofacial CT Scans to Stone-Cast Models Using 3D Symmetry for Backscattering Artifact Reduction*, Proceedings of Medical Image Understanding and Analysis Conference (MIUA) (2006); and A. Tariq and H. Foroosh, *Designing a Symmetric Classifier for Image Annotation using Multi-Layer Sparse Coding*, *Image and Vision Computing (IVC)*, Image and Vision Computing (IVC) (2018)

(initially published as *Image Annotation using Multi-Layer Sparse Coding*, at <https://arxiv.org/pdf/1705.02460> (May 6, 2017)).

9. Since joining UCF, I have advised and chaired the dissertation of over 30 doctoral students, 25 of whom have already successfully defended their doctoral dissertation, and secured academic positions or industry jobs in major high-tech corporations. As one example, I served as the advisor for a dissertation directed to “*Sub-pixel Registration in Computational Imaging and Applications to Enhancement of Maxillofacial CT Data*”. This research project focused on registration and alignment of 2D images and 3D digitized data obtained by dental CT scans and dental imprints of a patient (e.g., casts), with the aim of improving the quality of CT scans, e.g., eliminating backscattering issues. Methods were developed to register CT models to the physical imprints of patients (e.g., a stone cast model of the patient’s teeth).

10. I have been serving as the editor and/or committee member on numerous professional organizations. For instance, I have been involved with the following organizations: IEEE Transactions on Image Processing (Associate Editor, 2003-2008, 2011-2016); IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Member, Technical Committee, 2004-2012, 2015-2016, 2019-2020); IEEE Transactions on Pattern Analysis and Machine Intelligence (Reviewer, 1996-present).

11. For my work, I have been honored with numerous awards, including the *Piero Zamperoni Award*, International Association for Pattern Recognition (IAPR) (2004). I have also received several Best Paper awards, including “*The Best Scientific Paper Award*” in the IAPR International Conference on Pattern Recognition in 2008, “*The Best Paper Award*” in the IEEE International Conference on Image Processing (ICIP) in 2018, and the *Distinguished Researcher*, UCF College of Engineering and Computer Science (2004-2005, 2005-2006).

12. In sum, I have studied, taught, practiced, and conducted research in the area of computer vision and computer modeling of physical structures, including dental structures.

13. A copy of my *Curriculum Vitae*, attached as EX1003, contains further details on my education, experience, publications, and other qualifications to render an expert opinion in this matter.

II. SCOPE OF WORK

14. I understand that a petition is being filed with the United States Patent and Trademark Office for *Inter Partes* Review of U.S. Patent No. 11,049,248 to Salah et al. (“the ’248 patent,” EX1001), entitled “Method for Analyzing an Image of a Dental Arch.”

15. I have been retained by Align Technology, Inc. (“Align”) to offer an expert opinion on the validity of claims 1-2, 4, 7-8, and 13-16 of the ’248 patent assigned to Dental Monitoring SAS (“Dental Monitoring”). I receive \$650 per hour for my services. No part of my compensation is dependent on my opinions or on the outcome of this proceeding.

16. I have been specifically asked to provide my opinion on the unpatentability of claims 1-2, 4, 7-8, and 13-16 of the ’248 patent. In connection with this analysis, I have reviewed the ’248 patent. I have also reviewed and considered various other documents in arriving at my opinions and may cite to them in this declaration.

III. LEGAL STANDARDS

17. I understand that a claim is not patentable under 35 U.S.C. §102 as anticipated if the claim as a whole is described, either expressly or inherently, in a single prior art reference. I understand that this also requires that the prior art reference disclose that each claim element is arranged as recited in the claims.

18. I understand that differences between the prior art reference and a claimed invention, however slight, invoke the question of obviousness, not anticipation.

19. I have been advised that a claimed invention is not patentable under 35 U.S.C. §103 if it is obvious. A patent claim is unpatentable if the claimed

invention would have been obvious to a person of ordinary skill in the field at the time the claimed invention was made. This means that even if all of the requirements of the claim cannot be found in a single prior art reference that would anticipate the claim, a person of ordinary skill in the relevant field who knew about all this prior art would have come up with the claimed invention.

20. I have further been advised that the ultimate conclusion of whether a claim is obvious should be based upon several factual determinations. That is, a determination of obviousness requires inquiries into: (1) the level of ordinary skill in the field; (2) the scope and content of the prior art; (3) what difference, if any, existed between the claimed invention and the prior art; and (4) any secondary evidence bearing on obviousness.

21. I have been advised that, in determining the level of ordinary skill in the field that someone would have had at the time the claimed invention was made, I should consider: (1) the levels of education and experience of persons working in the field; (2) the types of problems encountered in the field; and (3) the sophistication of the technology.

22. I have also been advised that, in determining the scope and content of the prior art, in order to be considered as prior art, a reference must be reasonably related to the claimed invention of the patent. A reference is reasonably related if

it is in the same field as the claimed invention or is from another field to which a person of ordinary skill in the field would look to solve a known problem.

23. I have been advised that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was independently known in the prior art. In evaluating whether such a claim would have been obvious, I may consider whether there is a reason that would have prompted a person of ordinary skill in the field to combine the elements or concepts from the prior art in the same way as in the claimed invention.

24. I have been advised that any obviousness rationale for modifying or combining prior art must include a showing that a person of ordinary skill would have had a reasonable expectation of success.

25. I have been further advised that there is no single way to define the line between true inventiveness on the one hand (which is patentable) and the application of common sense and ordinary skill to solve a problem on the other hand (which is not patentable). For example, market forces or other design incentives may be what produced a change, rather than true inventiveness. I may consider whether the change was merely the predictable result of using prior art elements according to their known functions, or whether it was the result of true inventiveness. I may also consider whether there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the

patent. I may consider whether the innovation applies a known technique that had been used to improve a similar device or method in a similar way. I may also consider whether the claimed invention would have been obvious to try, meaning that the claimed innovation was one of a relatively small number of possible approaches to the problem with a reasonable expectation of success by those skilled in the art.

26. I understand that hindsight must not be used when opining on the patentability of claims of a patent. Therefore, my opinion will reflect the understanding of a person of skill in the art as of the relevant patent application filing dates, which I have been instructed is July 21, 2017, for this matter, the earliest date to which applicants have claimed priority for the claims of the '248 patent.

27. I have been instructed for claims to be entitled to an earlier application's filing date, the claims must find written description support in the earlier application to which benefit is claimed by an applicant.

28. I also understand that, in deciding whether the patent satisfies this written description requirement, one must consider the description from the viewpoint of a person skilled in the field of technology of the patent when the application was filed. The written description requirement is satisfied if a person having skill in the art reading the original patent application would have

recognized that it describes the full scope of the claimed invention as it is finally claimed in the publication that claims priority to the earlier application and that the inventor actually possessed that full scope by the filing date of the original application.

IV. BACKGROUND IN THE ART

29. The '248 patent describes methods involving computer vision and deep learning, with specific focus on analysis of dental images, such as images of the teeth of patients wearing an aligner. In the background below, I provide a brief overview of some important elements of computer vision and deep learning that are relevant to understanding what a POSA would have known in these areas at the time of filing of the '248 patent.

A. Image Analysis and Computer Vision

30. Images are a rich source of information that humans typically manage to analyze with relative ease. Image analysis is often described as the process of extracting information from images that may be subsequently used by either humans or computers to perform certain tasks. Developing methods and computer algorithms to provide the capability of extracting accurate and meaningful information from images has been the subject of study in computer vision in the past few decades.

31. Historically humans have been outperforming computers in tasks that require recognition and understanding of images in terms of their contents and the contextual relationships. It was not until after the early years of the last decade, with major breakthroughs in deep learning (both in terms of architecture and hardware support), that computers managed to also outperform humans in tasks such as recognition and image content and context understanding. *See* EX1015, 1-2, 8. Developments of parallel processing hardware (in particular Graphic Processing Units) made it possible for deep learning networks to be trained in an end-to-end manner and on much larger training sets, typically many orders of magnitude larger than before, involving tens of thousands to millions of images.

32. Computer vision is often described in layman's terms as the science of making computers see. One way to conceptually divide vision tasks, based on ideas pioneered by David Marr, divides such tasks into three levels: low-level, mid-level, and high-level tasks (also referred to as computation, algorithm, and implementation tasks). EX1016, 1538-39, 1541. Accordingly, computer vision algorithms were traditionally developed to handle low-level, mid-level, and high-level vision problems. Low-level computer vision refers to computational methods and algorithms that process images at pixel-level, modifying pixel properties to achieve tasks such as quantization, encoding, image restoration, image enhancement, etc. Mid-level computer vision refers to computational methods to

produce feature-level representations of images, such as color features, edge features, 3D geometric features, etc. A common feature-level representation widely used in computer vision tasks is boundary-based representations, such as edges and contours that define the separation between regions in images. Finally, high-level computer vision refers to computational methods and algorithms that aim to understand image contents and contextual relationships, including tasks like object classification, object localization, object instance recognition, etc.

33. Low-level and mid-level processing steps can be intermediate steps towards performing high-level vision tasks. For example, one may perform some restoration or enhancement (low-level) before extracting edges (mid-level) to align the edges with a template contour (low-level) to determine the presence and location of an object in the image (high-level), or to recognize an object or its class (high-level).

34. There were several motivating factors that led to using deep learning methods in computer vision, one of which was to overcome certain shortcomings, such as the treatment of different visual levels independently, as opposed to the way that humans integrate all levels of vision into a unified process. This led to a new class of computer vision methods that are designed to learn features from data, but also made it possible to take advantage of high-level labels to learn better features, while still relying also on low-level image properties and values to extract

features. This also made it possible for these methods to allow a network to learn all it needs in order to solve for a vision task, taking raw image data directly as input.

B. Edge Detection and Contour Detection

35. Edges, contours, boundaries, and discontinuities are various terms that have been used interchangeably in the literature to refer to similar concepts involving a representation of a discontinuity in an image. Edges and contours have long been considered as important features used for performing various high-level computer vision tasks, such as classification, object recognition, instance identification, etc. They have also been proven extremely useful for other low-level vision tasks, such as registration and alignment, or template matching that lead to high-level processing and annotation.

36. Classical edge detection was based on low-level properties, such as detected discontinuities, which would then be assigned a binary label (e.g., a 0 or 1) to identify the location of edges. Due to the key role of edges in computer vision and the rising influence of machine learning methods, researchers developed machine-learning based methods (including what would come to be referred to as deep learning) for edge and contour detection. EX1017, 1-2; EX1007, Abstract. Instead of binary labels (such as 1 and 0) for edge locations, deep-learning methods lead to multiple labels in the feature space (i.e. the contour image).

EX1017, 5-9; EX1007, 8, 9-12. What this means is that each contour is labeled differently based on various different high-level labels (e.g., class, instance, etc.) reflecting what is depicted within the region for which the contour is a boundary. For instance, contours of an object class A might be labeled as 1, those of an object class B as 2, those of an object class C as 3, and so on and so forth. These labels can be located on the contours of the object or assigned to the pixels in regions defined by the contour. EX1007, 9-12. Thus, deep-learning methods allow categorization of different objects in an image into specific classes.

37. These type of contours or boundaries are sometimes referred to as “category-specific” or “semantic” contours of boundaries. EX1017, 9; EX1007, 11. Thus, if a neural network is trained to look for features based on a training set with high-level labels (e.g., class labels), the trained machine would be able to return also features (e.g., edges) that are specific to that class. For instance, if a machine is trained to look for the contours of different types of teeth in intraoral images, it would not only detect edges (as in classical methods) but also would be able to label edges and the regions they define based on which type of tooth they belong to instead of a simple binary labeling.

C. Deep Learning

38. Deep learning refers to a class of machine learning methods that rely on many layers of feature learning in order to model the often-non-linear

relationship between the input (i.e. the training set) and the output (e.g., probabilities of classes or instances). These models are referred to as “deep” due to the fact that they consist of many layers between the input and the output, also known as the hidden layers, with these layers potentially comprising millions of parameters to be adjusted (i.e. to be trained) through a training process using a training dataset, in order to produce accurate outputs for the unseen test data.

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. A key characteristic of DNNs is the fact that the output of each hidden layer may be viewed as a feature map, with features being learned through a training process that yields an optimal feature map at the output of each hidden layer for the problem at hand. In other words, each layer of a trained model effectively functions as a feature extraction algorithm with features that are learned through a training process that upon convergence yields an optimal feature map at the output of each hidden layer for the problem at hand. Furthermore, as we go deeper into the network, the extracted features may be interpreted as higher level.

40. This important property of DNNs implies that the feature maps produced by earlier layers represent lower-level features, whereas the deeper ones correspond to higher-level features. This can be seen in Maninis, for example,

which shows how its DNN detects lower, level (fine-scale) features in early layers and higher-level (large-scale) features in later layers. *See* EX1007, 3 (“CNNs are by construction multi-scale feature extractors. If one examines the standard architecture of a CNN consisting of convolutional and spatial pooling layers, it becomes clear that as we move deeper, feature maps capture more global information....”). In applications involving images as input such as image classification or object recognition, 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 or overall object shape and class. *See* EX1007, Fig. 2. 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 that are optimized through a training process with large training datasets. *See* EX1007, 7-9 (“human” and “super-human” performance in various tasks).

V. THE '248 PATENT

41. The '248 patent is entitled “Method for Analyzing an Image of a Dental Arch.” EX1001, cover (54). The '248 patent describes acquiring dental images using an image acquisition apparatus, which may be a cellphone. EX1001, Abstract, 8:47-53. Independent claim 1 recites as follows:

1. A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a") more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image", the analysis image being a photograph, or an image extracted from a film;

b") analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for at least one tooth attribute of an "analysis tooth zone" representing, at least partially, a tooth on said analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image,

in the step a"), a cellphone is used to acquire the analysis image.

EX1001, 32:6-23. Summarized briefly, claim 1 includes steps of acquiring, with a cellphone, an image of a patient's teeth and an aligner worn in a service position on the teeth; analyzing the image with a trained deep learning device; and determining a value for a tooth attribute of a zone at least partially representing a tooth on the image. The tooth attribute must be an attribute relating to a separation between the tooth and the aligner.

42. The other independent claim, claim 15, includes language that overlaps substantially with claim 1. Claim 15 recites as follows:

15. A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a") more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image" the analysis image being a photograph, or an image extracted from a film;

b") analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for an image attribute of the analysis image, the image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on said analysis image.

EX1001, 34:43-57. The preamble and step (a") of claim 15 are identical to claim 1; however, claim 15 does not include the limitation at the end of claim 1 requiring that a cellphone is used in step (a"). Claim 15 also differs in that rather than determining a tooth attribute, it is an image attribute that is determined.

43. Dependent claims 2 and 16 recite that the aligner's suitability is assessed based on the tooth or image attribute. EX1001, 32:24-27, 34:58-61.

Dependent claim 4 recites sending an information message as a function of the assessment in claim 2. EX1001, 32:33-36. Dependent claims 7 and 8 recite

limitations regarding training of the deep learning device with images showing tooth or image attributes. EX1001, 32:45-33:15. Claim 14 recites limitations

similar to claim 7, and it further specifies that probabilities are determined relating

to the presence of an analysis tooth zone and a tooth attribute for the corresponding tooth. EX1001, 34:21-42. Finally, claim 13 recites that the deep learning device is a neural network. EX1001, 34:19-20. As I explain in further detail below in section IX, these dependent claims merely add routine aspects known in the art and taught or suggested by the grounds references.

VI. PRIOR ART

A. Kim

44. U.S. Patent Publication No. 2017/0281313 (“Kim,” EX1005) is entitled “Method and System for Providing Remote Teeth Alignment Services and Computer-readable Medium on which the Method is Recorded.” Kim was filed on April 5, 2016, and published on October 5, 2017. I understand that Kim may be considered in evaluating the patentability of the claims of the ’248 patent.

45. Kim teaches methods and systems to provide “remote teeth alignment services.” EX1005, Abstract. Kim teaches methods by which an orthodontist can be provided “accurate teeth alignment information using a device, e.g., a smartphone, of an orthodontic patient,” including information allowing the orthodontist to “diagnose the condition of the patient based on the teeth alignment information” transmitted as well as to “remotely provide information about an accurate step of a clear aligner to be currently worn by the orthodontic patient.” EX1005, ¶10. Kim refers to the patient’s smartphone as an “orthodontic patient

terminal,” though Kim notes that a variety of other patient devices, such as mobile devices, laptop computers, smart cameras or camcorders, and similar devices may also be used. EX1005, ¶¶41-42.

46. Kim’s teaches obtaining teeth alignment information by taking a variety of photographs of the patient’s teeth from different angles. EX1005, ¶¶55-56. Additionally, Kim teaches that the patient’s smartphone can send “clear aligner information about a currently worn clear aligner” to the orthodontist. EX1005, ¶66. This information is transmitted “in such a manner that the orthodontist checks a photo or a video of the clear aligner currently worn by the orthodontic patient, and determines whether the clear aligner is normal and corresponds to an appropriate step.” *Id.*

B. Invisalign

47. The Invisalign Clinical Monitoring Guide (“Invisalign,” EX1006) was published in 2003. EX1006, i. Invisalign was archived by the Internet Archive at http://web.archive.org/web/20030926024959/http://www.invisaligncec.com:80/consistent/pdfs/ICMG_Final.pdf, reflecting an archive date of September 26, 2003, indicating it was published no later than that date. I understand that Invisalign may be considered in evaluating the patentability of the claims of the ’248 patent.

48. Invisalign teaches methods for monitoring progress of orthodontic treatment when using clear plastic aligners manufactured by Align Technology,

Inc. EX1006, i. The guide describes how orthodontic patients' treatment progress can be monitored after the patient has received aligners and the aligners are being worn by the patient. *Id.* Invisalign identifies key considerations for effective use of aligners in treating patients, and it also describes the need to use high-quality records, including both dental impressions and photographs of the teeth of the patient wearing the aligners. EX1006, ii-iv.

49. Invisalign also provides specific examples of measurements to perform to track whether aligners are properly treating patients, as well as discussing what those measurements indicate about whether the aligner is appropriate in terms of fit, seating, and other properties. *See, e.g.*, EX1006, 1-2. In one example, Invisalign illustrates how poor aligner seating may be assessed, showing with a digital photograph of an aligner on a patient's teeth how a poor fit can be determined. *See* EX1006, 1. In the example photograph, the poor fit is shown by the presence of a gap of over 1 mm between the ends of the teeth and the corresponding parts of the aligner that receive each tooth. *Id.* This photograph is shown below; the image in Invisalign includes overlaid arrows pointing out the gaps that are visible for multiple teeth. *Id.*



C. Salah

50. PCT Publication No. WO2016066651 to Salah et al. (“Salah,” EX1008) is entitled “Method for Monitoring Dentition.” Salah was published in French, but I understand that EX1008 is a certified translation of Salah. Salah names 4 inventors in common with the ’248 patent. Salah was published on May 6, 2016, more than 1 year before the earliest claimed priority date of the ’248 patent. Accordingly, I understand that Salah may be considered in evaluating the patentability of the claims of the ’248 patent.

51. Salah discloses methods for monitoring patients’ teeth, which can be helpful to track the teeth during orthodontic treatment. EX1008, Abstract, 1:12-22, 8:10-18. Two-dimensional images can be obtained of a patient’s teeth, and these

images can be compared to a three-dimensional model of the patient's teeth, allowing the model to be updated and used to measure progress of treatment. EX1008, 1:26-2:19, 46:15-19, Fig. 1. Salah's analysis of images can include using contour detection on an acquired image to aid mapping to a reference model, and this contour detection can be performed using deep learning. EX1008, 34:11-14, 36:14. Although Salah uses the term "deeplerning," rather than deep learning, this appears to be a typographical error; a POSA would have understood that Salah meant to be referring to "deep learning," as deep learning is a well-known method of computer analysis that has many uses, including specifically the detection of contours in images. EX1008, 36:14.

52. Salah teaches that its 2D images may be obtained by a patient or a relative using a device such as a cell phone, which can also perform the remaining steps of the evaluation process (including comparison to a 3D model). EX1008, 2:21-27, 31:4-7. Because the users may not have training in image acquisition or dentistry, Salah also teaches that the image-acquisition apparatus can guide the user to obtain appropriate images. EX1008, 27:21-28:10. For example, the apparatus can indicate whether the image is acceptable or needs to be retaken, and it can also provide indications (such as arrows or other instructions) indicating how to adjust the camera's positioning to correctly orient it to obtain an image. *Id.*

D. Maninis

53. 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. Copies of Maninis are available to download from links at an abstract page (<https://arxiv.org/abs/1701.04658>), with two versions available: version 2 (EX1007), dated April 28, 2017, and version 1 (EX1021), dated January 17, 2017. The download link for version 2 is <https://arxiv.org/pdf/1701.04658> (or <https://arxiv.org/pdf/1701.04658v2>), and the download link for version 1 is <https://arxiv.org/pdf/1701.04658v1>. The version submitted as EX1007 was downloaded from the version 2 link. The two versions of the paper contain only minor differences, but for purposes of this declaration, I focus on the April 28, 2017, version of Maninis (EX1007).

54. I am familiar with arXiv.org as a user of the system and a registered author of numerous articles available from the system. In May 2017, I submitted several papers to arXiv.org in the same field, so I am familiar with the processes practiced by arXiv.org for publishing submitted papers at the time that Maninis was submitted. arXiv.org is an openly accessible, moderated repository for scholarly articles, covering an array of scientific fields. Authors submit their articles to arXiv.org to disseminate them freely to interested members of the public. Authors of scientific papers use arXiv.org as a method of publishing their

papers that reliably establishes a date of publication, as the arXiv.org repository permanently marks each document with a stamp indicating the date of submission, and arXiv.org consistently publishes submitted documents to the interested scientific community promptly after submission, as I discuss below.

55. The date of submission of a document is reflected in a stamp on the left side of the paper's first page; for example, Maninis indicates a submission date of April 28, 2017. EX1007, 1; EX1022. Authors indicate the field relevant to a paper at the time of submission, which is reflected as part of a stamp on the paper's first page. For example, the stamp for Maninis includes the code "[cs.CV]," indicating that the paper is a computer science paper in the field of computer vision. EX1007, 1.

56. As shown on an archived webpage of the arXiv.org website, it was the established policy of arXiv.org in 2017 to publish submitted papers within about one business day of the date of submission. EX1022 (archived May 3, 2017). As the arXiv.org website explains, a submitted paper will be announced about one day after being submitted, and it will then be mailed to all arXiv.org subscribers (who would typically be persons interested in the paper's relevant field) shortly thereafter. EX1022. For example, a paper submitted on a Friday would be announced Sunday or Monday and mailed to subscribers no later than Tuesday morning. EX1022. This is consistent with my experience with arXiv.org typical

practice in early 2017, as I observed regarding papers I submitted and papers I received as a subscriber during that time. Regarding Maninis, with a submission date of April 28, 2017, it would have been announced on Sunday April 30, 2017, or Monday May 1, 2017, and would have been transmitted to subscribers no later than Tuesday May 2, 2017.

57. Once a paper is announced, it was arXiv.org's stated policy that the paper may not be modified without creating a new version. EX1023 (archived July 18, 2017); EX1026 (archived September 8, 2016). Maninis itself reflects this practice, as the document indicates that it is the second version (with v2 as part of the stamp on the left side of the document), meaning that it was a revision made on April 28, 2017, to an earlier version (that had been submitted on January 17, 2017, *see* EX1021). Thus, under arXiv.org's standard practice, the version of Maninis submitted on April 28, 2017, would not have been changed after its announcement (on or about April 30, 2017). I have reviewed metadata contained in a copy of Maninis downloaded from the download link (<https://arxiv.org/pdf/1701.04658>), and this metadata indicates that the document was created and last modified on April 30, 2017. This is consistent with Maninis being placed in final form by arXiv.org as of an announcement date of April 30, 2017, with the version of Maninis that was published at that time being the same as the document provided

as EX1007. This is also consistent with my experience with arXiv.org regular practice from early 2017 until today.

58. The public availability of Maninis to interested persons prior to July 21, 2017, is also corroborated by archived webpages of both arXiv.org and the authors of Maninis. As of May 3, 2017, the university website featuring the author's Convolutional Oriented Boundaries software included a link (to <https://arxiv.org/abs/1701.04658>) for downloading a PDF copy of Maninis from arXiv.org. EX1024 (archived May 3, 2017). Additionally, as of July 8, 2017, the arXiv.org author search webpage for Prof. Luc Van Gool, the last author listed on Maninis, listed Maninis as one of his recent papers that was available on arXiv.org, providing links to both the abstract page (<https://arxiv.org/abs/1701.04658>) and a link to directly download Maninis (<https://arxiv.org/pdf/1701.04658>). EX1025 (archived July 8, 2017). Thus, Maninis was available openly to general public and disseminated to the scientific community prior to July 21, 2017. Accordingly, I understand that Maninis may be considered in evaluating the patentability of the claims of the '248 patent.

59. 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 that define regions containing objects. EX1007, 3-4, 5-6, 13, Fig. 2. Maninis accomplishes this contour detection by using a deep learning neural network to generate a “contour probability map,” which is transformed “into a hierarchical boundary map,” with different probabilities translating to different “strengths” of borders between regions containing objects in the image. EX1007, 3, 5. Maninis also demonstrates the ability to label the specific object contained within each identified region, using a process called semantic boundaries, which Maninis shows to also improve over the state of the art. EX1007, 11-12, Tables 4-7, Fig. 16.

VII. OVERVIEW OF THE LEVEL OF SKILL

60. I have been advised that “a person of ordinary skill in the relevant field” (or “person of ordinary skill in the art,” or “POSA”) is a hypothetical person to whom one could assign a routine task with reasonable confidence that the task would be successfully carried out. I have been advised further that, in determining the level of ordinary skill in the art that someone would have had at the time the claimed invention was made, I should consider: (1) the levels of education and experience of persons working in the field; (2) the types of problems encountered in the field; and (3) the sophistication of the technology.

61. In my opinion, a person of ordinary skill in the relevant field as of July 21, 2017, would typically have an advanced degree, such as a Masters’

Degree or Ph.D. in computer science or a related field, with at least two years of work in computer vision, image analysis, medical imaging, or related areas. A POSA could have less education but significant professional experience in one or more of these fields. A POSA could be a member of an interdisciplinary team including persons with backgrounds such as electrical engineering, computer science, optical imaging, and dentistry or orthodontics, and could call upon the knowledge of other team members as appropriate.

VIII. CLAIM CONSTRUCTION

62. I have been advised that the claim terms of the '248 patent claim are to be given the plain and ordinary meaning that the terms would have to a person of ordinary skill in the art in question at the relevant priority date. I understand that the analysis focuses on how the patentee used the claim term in the claims, specification, and prosecution history. I also understand that dictionaries or other extrinsic sources may assist in determining the plain and ordinary meaning but cannot override a meaning that is unambiguous from the intrinsic evidence. I have followed these principles in my analysis throughout this declaration.

63. The scope of the challenged claims is discussed throughout my analysis; I provide additional discussion of some claim terms below.

A. tooth attribute

64. Claim 1 of the '248 patent recites a “tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image.” The specification of the '248 patent distinguishes between image attributes 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. However, the specification also explains that a tooth attribute may 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, 28:48-55. The '248 patent states that one or more tooth attributes can be “associated with the tooth models as a function of the teeth that they model,” as well as referring to “a tooth attribute relating to a tooth represented on the analysis image.” EX1001, 1:26-27, 70:62-63. The specification also lists non-exhaustive examples of tooth attribute categories. *See* EX1001, 7:64-8-42. For example, the specification states that

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

example “abraded”, “broken”, “decayed” or “fitted” (that is to say in contact with a dental, preferably orthodontic, appliance), an age for the patient, or a combination of these attributes. A tooth attribute is preferably an attribute which only relates to the tooth modeled by the tooth model.

EX1001, 7:64-8-9; *see also id.*, 32:40-44 (providing similar list for claim 6). This description confirms that the term “tooth attribute,” on its own, refers to any attribute related to a tooth.

65. In the particular context of the claims, the tooth attribute is specified to be an attribute “relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image.” The specification gives some examples of tooth attributes of this type, stating that the attribute may be “chosen from the group formed by a maximal separation along the free edge of the tooth [or] a mean separation along the free edge of the tooth.”

EX1001, 28:48-55. However, the specification merely indicates that these options are preferred examples, not an exhaustive list. *Id.* Indeed, dependent claim 6 indicates that claim 1’s “tooth attribute relating to a separation...” encompasses a broader realm of attributes, not merely attributes involving a quantification of the separation. Claim 6 indicates that such a tooth attribute can be 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.” EX1001, 32:40-44. This confirms that the “relating” recited in claim 1 is not to be interpreted narrowly to require the tooth attribute to be a value or measurement of separation. Rather, the recited tooth attribute may be related to the separation more indirectly—for example, a “tooth shape parameter” would not directly provide a measure of separation from an aligner, but it might “relate” to a separation in that the shape of the tooth includes the shape of the edge of the tooth facing the aligner, and it is from such an edge that separation from the aligner may be measured. Accordingly, the “tooth attribute” recited in claim 1 includes any attribute of a tooth that may relate to the tooth’s separation from an aligner, including attributes specific to the tooth or to the aligner, such as their respective positions or edges, which may bear a relation to the separation between the tooth and the aligner.

B. image attribute

66. Claims 8 and 15 recite limitations relating to an “image attribute.” In particular, claim 8 recites an “image attribute relating to a separation between the tooth or teeth and the aligner represented on the historical image” and claim 15 recites an “image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on said analysis image.” The ’248 patent describes an image attribute as an attribute that has a value that “relates to the whole image and not to part of the image.” EX1001, 3:23-

37. The '248 patent lists various examples of what an image attribute may relate to, including attributes relating 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. Regarding the “value” of an image attribute, the specification indicates that the value need not be a single numerical value. For example, a list of values for a set of teeth would also be an image attribute, such as a list of all tooth numbers of teeth represented in an image. EX1001, 26:3-14. These various examples indicate that 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.

67. Claims 8 and 15 recite that the image attribute relates to a separation between one or more teeth and the aligner represented on either a historical image (for claim 8) or an analysis image (for claim 15). In some examples, the '248 patent states that the image attribute may be “chosen from the group formed by a

maximal separation along all of the teeth represented, a mean separation along the free edges of all of the teeth represented, an overall acceptability of the separation of the teeth represented,” but it indicates that these are preferred options, not an exhaustive list. EX1001, 28:48-55. The ’248 patent also teaches that an image attribute may be “determined as a function of the tooth attribute values.” EX1001, 26:3-14. As I discussed above, an image attribute may be a single value or a list of values for multiple teeth, such as a “list of the tooth numbers represented” in an image. *Id.* As a further example, an image attribute may also indicate “the ‘presence’ or ‘absence’ of a dental, preferably orthodontic, appliance.” EX1001, 26:11-14; *see also id.*, 29:30-53 (determination of image attribute can be “determination...of the presence of a separation between the aligner and the tooth or teeth represented on the analysis image, and/or of an amplitude of said separation”). In view of the ’248 patent’s description, and similar to the tooth attribute “relating to” a separation in claim 1, the “image attribute” recited in claims 8 and 15 includes any attribute of the image as a whole that relates to the separation between one or more teeth and the aligner. For example, among these attributes would be an indication of whether an aligner or gap was present in the image, a measure of average separation, a measure of separation for one or more teeth, an indication of whether the overall separation is acceptable, or a measure of another value that may be used to assess separation of teeth from the aligner.

C. image extracted from a film

68. Claims 1 and 15 require an analysis image that is “a photograph, or an image extracted from a film,” and claim 1 also specifies that the analysis image is acquired with a cellphone. The specification states that a patient “preferably takes a photo or video of his or her teeth, preferably by means of a cellphone equipped with a camera,” and elsewhere it states that the “analysis image is preferably a photo or an image from a film, preferably with a cellphone.” EX1001, 19:41-44, 25:54-57. In light of this parallel disclosure and the claims’ requirement that the analysis image be acquired with a cellphone, the phrase “image extracted from a film” refers to an image extracted from a video recording.

IX. CLAIMS 1-2, 4, 7-8, AND 13-16 WERE OBVIOUS AS ILLUSTRATED BY KIM, INVISALIGN, SALAH, AND MANINIS

A. The Combination of Kim, Invisalign, Salah, and Maninis

69. Kim discloses methods for providing remote tooth alignment services, including evaluation of a clear aligner to be worn by a patient and determination of, for example, “when a new or modified clear aligner is necessary.” EX1005, Abstract, ¶10. Kim describes providing these services using an “orthodontic patient terminal,” which may be a patient’s smartphone. EX1005, Abstract, ¶¶1, 10. 42. To facilitate these services, photographic or video images may be transmitted to an orthodontist to provide information about teeth alignment and the patient’s clear aligner. EX1005, ¶12. Kim also teaches that automatic recognition of contours of a

patient's face and teeth may be employed as part of the imaging process. EX1005, ¶14. As one example, Kim teaches providing a "teeth contour automatic-recognition photographing program for displaying a virtual teeth line on the screen of the orthodontic patient terminal and automatically recognizing and taking a photo of teeth of the orthodontic patient shown on the screen." *Id.*

70. Kim further describes how a "clear aligner information transmission program" may operate to provide orthodontists with images of aligners worn by the patient that may be usefully analyzed by the orthodontist. EX1005, ¶66. The program receives "clear aligner information about a currently worn clear aligner from the orthodontic patient terminal." *Id.* This clear aligner information can include a photo or a video which is stored and sent to the orthodontist. *Id.* Then, "the orthodontist checks a photo or a video of the clear aligner currently worn by the orthodontic patient, and determines whether the clear aligner is normal and corresponds to an appropriate step." EX1005, ¶66. Kim teaches that its techniques provide reduced diagnosis time and cost relative to traditional in-office visits, as well as providing for enhanced accuracy. *See* EX1005, ¶¶8-10, 72. For example, Kim allows remote monitoring to substitute for traditional visits to orthodontists' offices (*see* EX1005, ¶¶8-9), by allowing an orthodontist to "obtain accurate teeth alignment information using a device, e.g., a smartphone, of an orthodontic

patient,” allowing remote diagnosis, as well as increasing accuracy of diagnosis using images acquired of patients’ mouths. EX1005, ¶10.

71. Invisalign corroborates Kim’s teachings regarding imaging an aligner in a service position on teeth, showing how such an aligner may be assessed to determine if it fits appropriately. Moreover, Invisalign’s teachings make it readily apparent to do so in a system such as Kim’s. Invisalign specifically details assessments that can be used to determine whether the clear aligner is normal and corresponds to an appropriate step, teaching how to “successfully monitor Invisalign treatments and address issues that may arise between receipt of Aligners from Align to the completion of treatment.” EX1006, i. The first criterion identified by Invisalign addresses “General fit issues,” with Invisalign specifying that if an aligner is well seated, there will be a distance of less than 1 mm between the end of the teeth and the surface of the aligner. EX1006, 1. This type of separation is depicted in a digital photograph, where a poorly-seated aligner is shown being worn by a patient in a service position, with arrows indicating that the aligner surfaces corresponding to teeth (i.e., the parts of the aligner’s tooth-shaped surfaces that match with the end of each given tooth) have a separation of greater than 1 mm from the ends of the teeth. *See id.* The resulting gap is plainly visible on the photograph. Describing the gap in terms appropriate for a computer-vision application, the gap corresponds to the space between the contour defining the

occlusal end of the teeth and the matching contour defining the occlusal surface of the clear aligner. The photograph is reproduced below, and the shapes of each surface can be naturally perceived by the human eye, with the edge of the teeth visible as a transition from white to black and the edge of the aligner being what the arrows point towards:



EX1006, 1. Thus, Invisalign teaches that orthodontic assessment of an aligner may include assessing the separation between the edges of the aligner and the patient's teeth when the aligner is worn, as may be seen in a digital photograph showing the patient's teeth as well as the clear aligner being worn in a service position on the patient's teeth.

72. In view of Kim's disclosure of orthodontic checking of a photo or video of an aligner currently worn by the patient and Invisalign's teachings of how to evaluate aligner fit, illustrated by an image of an aligner worn on a patient's teeth, a POSA would have had reason to analyze images of the teeth of a patient wearing an aligner in a service position to evaluate whether the aligner fits appropriately. This combination would be applying known methods for orthodontic evaluation of aligner fit based upon the appearance of the aligner worn on the patient's teeth, as taught by Invisalign, in order to analyze a photo or video of the aligner worn by the patient, as taught by Kim, thereby providing feedback regarding tooth-aligner separation to aid an orthodontist regarding whether the clear aligner is normal and corresponds to an appropriate step of treatment. As Kim indicates, employing such techniques would serve multiple goals. As one example, Kim teaches that accuracy would be enhanced. *See* EX1005, ¶¶10, 72. Kim also teaches that the time and cost of treatment may be reduced. *Id.* Further, implementing assessment as taught by Invisalign and Kim would facilitate remote assessment, reducing the burden of office visits for orthodontics patients. *See* EX1005, ¶¶8-10. Additionally, a POSA would have understood that the type of image analysis suggested by Kim and Invisalign could be readily applied on images or videos showing a clear orthodontic aligner being worn on the teeth of a patient. For example, a POSA would have recognized that by using computer-

vision techniques to detect the locations of the ends of the patient's teeth and the corresponding surfaces of the aligner, the distance between the two could be measured, corresponding to the assessment taught by Invisalign.

73. Not only are these types of images of aligners on teeth expressly depicted in Invisalign (*see* EX1006, 1), but the fact that POSAs were motivated to use such images for orthodontic evaluation was confirmed in publications by practicing orthodontists. For example, Dr. Scott Frey published a step-by-step guide on the Ortho Cosmos website explaining how to conduct a simple remote orthodontic monitoring system allowing orthodontists to analyze orthodontic patients' teeth and aligners over the internet. EX1009, 17-27. The guide explains that patients would be shown example photos or videos to show exactly what images they should capture, including one video "with teeth apart and aligners in." EX1009, 19. Patients could then record and transmit such videos with their smartphone, providing orthodontists with video images sufficient to conduct typical orthodontic treatment evaluations. EX1009, 20-27. Thus, published descriptions show that photographs or videos of the teeth of patients wearing aligners, including images taken with cellphones, were already known to be useful to analyze for orthodontic assessment purposes.

74. A POSA would have known how to determine the width of such gaps using image processing software. As I discussed briefly above, the distance

between depicted teeth and a depicted aligner corresponds to the distance from the edge of one to the edge of the other. Salah provides disclosure useful for determining such edges. Like Kim, Salah teaches application of computer image analysis to dental images for orthodontic treatment purposes. Salah teaches methods that are useful for monitoring patient's teeth, for example, to track the teeth during orthodontic treatment, including steps of acquiring images of the patient's teeth at various times during treatment. EX1008, Abstract, 8:19-28. Salah teaches that the acquired images can include two-dimensional images, which may be acquired by an image acquisition apparatus such as a mobile phone. EX1008, 13:24-28, 31:4-9. Further paralleling Kim, Salah teaches that images may be analyzed to "measure the effectiveness of orthodontic treatment," including by "remotely monitor[ing] the progress of an orthodontic treatment." EX1008, 46:15-24. Salah teaches that analysis of acquired images may include determining contours of objects in an image, including by using deep learning for contour detection. EX1008, 34:5-36:14.

75. To the extent that Salah does not expressly teach using a deep learning device trained by means of a learning base for image analysis, this would have been obvious to a POSA and is taught by Maninis. Maninis teaches a method referred to as "Convolutional Oriented Boundaries (COB)," which employs a deep learning device to perform image classification, object detection, and semantic

segmentation. EX1007, Abstract, 11-12. Maninis provides methods of “contour detection” that produce “multiscale oriented contours and region hierarchies” using “Convolutional Neural Networks (CNNs).” *Id.* To accomplish this, Maninis’s device divides images into regions containing different objects, in both foreground and background, and then classifies the objects into appropriate categories. *See* EX1007, 1-2, 5-6, 8, 11-12. Maninis accomplished these tasks by training its neural network on an appropriately-labeled training dataset—that is, a “learning base” in the terminology of the ’248 patent. EX1007, 1-2, 7-8.

76. Although Maninis is not restricted to dental imaging, Maninis describes broadly applicable techniques for image classification and object detection using category-specific contour detection based on a deep learning device trained using a learning base. For example, Maninis describes the ability of its system to transfer between different types of neural network architectures and thereby “transfer further improvements ... to perceptual grouping.” EX1007, 13. Maninis also confirms that its process is not narrowly limited to fixed categories, and can in fact generalize concepts of objects even to categories it has not originally seen. EX1007, 10-11. Salah teaches that image analysis techniques for detecting contours may employ deep learning, which would motivate a POSA to apply Maninis’s teachings to the imaging apparatus of Kim and Salah. EX1008, 34:5-36:14. Although Salah notes that certain other image analysis techniques are

“preferred,” it notes that even if not preferred, deep learning “can also be used.” *Id.* A POSA would not interpret Salah’s identification of some techniques as preferred over deep learning as a reason not to apply deep learning; rather, a POSA would have recognized that Salah had identified deep learning as a useful option that can be applied. Moreover, as I explain below (as well as in section IV.C), trends developing in computer vision at the time favored increased adoption of deep learning for image analysis, which would further motivate a POSA to employ it in applications like those disclosed by Salah. Indeed, Maninis itself explains that the use of deep learning for object and contour detection represented the established trend in computer vision, confirming that Maninis would be a natural selection for such a technique. EX1007, 2.

77. A POSA would view Maninis as providing good reasons to employ 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 explains that its COB technique is compatible with different CNNs and is applicable in a “modular” fashion that “can incorporate and benefit from future improvements in the base CNN.” *Id.* Maninis describes that its technique provides improvements relative to “recent approaches to contour detection using deep learning.” EX1007, 3, 10. For example, Maninis demonstrates substantially faster object edge and contour detection and image

classification relative to state-of-the-art techniques. *See* EX1007, 10 (“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 (describing improvements in semantic segmentation to classify specific types of depicted objects).

78. Maninis’s flexibility and prospect to improve performance would motivate a POSA to analyze images using a deep learning device, as taught by Maninis, to improve Kim and Salah’s image acquisition and analysis procedures. For example, a POSA would have been motivated to adapt Salah and Kim’s image acquisition and analysis devices to employ a deep learning architecture as taught by Maninis, trained on images relevant to the dental imaging applications of Salah, Kim, and Invisalign. In that combination, Maninis’s deep learning architecture would provide specific functionalities required by Salah, Kim, and Invisalign, including the ability to identify objects (such as the teeth and aligners depicted in each image) and to detect contours (such as the edges of teeth and aligners useful to assess aligner fit as taught by Invisalign).

79. A POSA would have had reason to use the techniques of Salah and Maninis for contour detection to detect the edges of teeth and aligners, as Invisalign teaches that the gap between the surfaces of teeth and an aligner should be below a threshold value (1 mm) when fitting well, aligning with Kim’s teaching

to assess whether an aligner was “normal and corresponds to an appropriate step.” EX1005, ¶66; EX1006, 1. As I explained above, the width of a gap between two depicted objects in an image (such as a tooth end and a surface of an aligner) corresponds to the distance between the contours for each object. Equivalently, the “gap” is a portion of the image between those surfaces, and its width is the distance from one side to the other (i.e., the distance from a contour separating “tooth” from “gap” and a contour separating “gap” from “aligner.”). Thus, to apply Invisalign’s teaching in the context of Kim, Salah, and Maninis, a POSA would have analyzed an image of an aligner worn in a service position on teeth by measuring the distances from the contour defining the ends of the teeth to the contour defining the surfaces of the clear plastic aligner, as detected by a deep learning device.

80. A POSA would have reasonably expected such a combination to work predictably and successfully. 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 image acquisition and analysis apparatus as taught by Kim in view of Invisalign and Salah would also be consistent with trends in computer vision, as I describe above in section IV. Maninis agrees, noting 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. In combination, the

elements taught by each reference would perform substantially the same function as they provide alone, applying the contour detection taught by Salah and Maninis to detect edges of teeth and aligners to produce a measurement of distance between these surfaces, as taught by Invisalign, to assess whether the imaged aligner fits appropriately for the current step in patient treatment, as taught by Kim. Thus, based on Maninis's promising disclosure and reported improvements on the state-of-the-art, a POSA would have been motivated to incorporate a trained neural network as taught by Maninis to detect edges and identify objects in an image to help assess aligner fit based on separation between teeth and aligner, as taught and suggested by Kim in view of Invisalign, and Salah, with a reasonable expectation of success.

81. Below, I note specific prior art disclosure that satisfies each element of each challenged claim.

B. Claim 1

82. Kim, Invisalign, Salah, and Maninis teach each and every element of claim 1, as I explain below.

1. Claim 1 preamble

83. The preamble of claim 1 recites "[a] method for assessing the shape of an orthodontic aligner, said method comprising the following steps." Kim and Invisalign teach the preamble of claim 1.

84. Kim teaches evaluating an aligner by reviewing an image of the aligner currently worn by the patient, to determine whether the aligner is normal and appropriate. EX1005, ¶66. Invisalign teaches that evaluating a currently-worn clear aligner includes assessing the fit of the aligner, and that this assessment involves determining a distance between the teeth and the aligner when worn by the patient, as depicted in photographs showing good and poor aligner seating. EX1006, 1. Kim and Invisalign also teach that the evaluation can determine that an aligner does not fit appropriately (for example, because the aligner may have been formed using a distorted impression of teeth). EX1006, 1; EX1005, ¶¶66-67. In view of Kim and Invisalign, a POSA would have understood that the shape of an orthodontic aligner can be assessed by analyzing images of the aligner worn on a patient's teeth, including by Invisalign's teachings to determine whether the image shows that the aligner is shaped appropriately to fit the patient's teeth based on their depiction in one or more images acquired by Kim's smartphone.

85. Kim and Invisalign thus teach the preamble of claim 1.

2. Step 1(a'')

86. Step (a'') of claim 1 recites as follows:

a'') more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called

“analysis image”, the analysis image being a photograph, or an image extracted from a film;

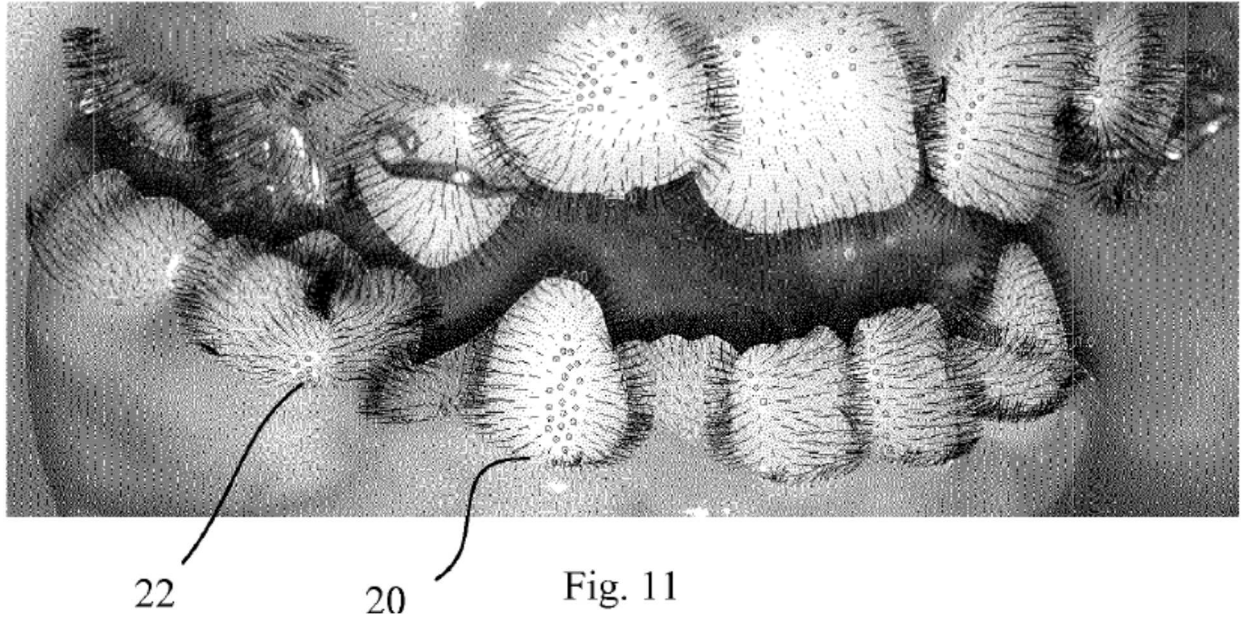
The last limitation of claim 1 further recites “in the step a”), a cellphone is used to acquire the analysis image.” Because this last limitation further characterizes step (a”), I address these limitations together as making up step (a”) of claim 1. Read together, step 1(a”) requires acquiring at least one image (the “analysis image”) showing a patient’s teeth and an aligner worn by the patient in a service position. The image must be a photograph or an image extracted from a film (i.e., from a video), it must be acquired more than 1 week after starting treatment with the aligner, and a cellphone must be used to acquire the analysis image. As I discuss below, Kim, Invisalign, and Salah teach these elements.

87. Kim teaches that a patient’s smartphone or similar device may be used to acquire orthodontic treatment information. *See* EX1005, Abstract, ¶¶1 (“smartphone of an orthodontic patient”), 10 (“[A]n orthodontist may obtain accurate teeth alignment information using a device, e.g., a smartphone, of an orthodontic patient....”), 42 (“Herein, the orthodontic patient terminal 10 is not limited to a smartphone and may be one of all information devices capable of transmitting a teeth alignment information.”). This can include “using a camera” of the device to acquire photographs “showing a teeth alignment state of the orthodontic patient.” EX1005, ¶¶55-56. This can further include aligner information, and in particular may include “receiving the clear aligner information

about a currently worn clear aligner,” enabling the orthodontist to “check[] a photo or a video of the clear aligner currently worn by the orthodontic patient.” Using this photo or video from the patient’s cellphone, the orthodontist “determines whether the clear aligner is normal and corresponds to an appropriate step” EX1005, ¶66. Thus, Kim teaches acquiring an analysis image of a patient’s teeth wearing an aligner as recited in step 1(a”).

88. Invisalign confirms that analyzing an image of an aligner in a service position on the patient’s teeth (such as a photo or video taken by a cellphone, as taught by Kim) can produce measurements useful for orthodontic assessment. Invisalign teaches that a large gap between a tooth and an aligner, such as a gap of more than 1 mm, indicates poor fit. EX1006, 1. Invisalign further illustrates that such a gap is visually distinguishable in a digital photograph of a patient’s teeth that at least partially represents an aligner worn in a service position on the teeth. *See id.* Accordingly, Invisalign provides a reason to acquire an image of the patient’s teeth with an aligner being worn in a service position: by assessing the gap visible between aligner and teeth, Invisalign teaches that it can be determined whether an aligner fits correctly (e.g., is normal and corresponds to an appropriate step), or whether it is seated or fit poorly. This assessment involves determining the size of the gap, with gaps above a given threshold (e.g., 1 mm) indicating a poor fit.

89. Salah similarly teaches acquiring dental images with a cellphone, including of teeth fitted with orthodontic appliances. For example, Salah teaches that a user—which may be a non-specialist such as a patient or the patient’s relative—may acquire images of a patient’s dentition using an image acquisition apparatus such as a mobile phone. EX1008, 31:4-7. The mobile phone can include a photographic or video camera. EX1008, 27:11-13 (“The image acquisition apparatus preferably comprises a camera or a video or infrared camera, which the user, e.g. the patient or one of their relatives, positions by means of a viewfinder or screen, before operating it”). Salah also teaches that the image of the teeth may include an orthodontic appliance; for example, Fig. 11 (below) depicts a reference map generated from an image of teeth of a patient wearing an orthodontic appliance. EX1008, 64:7-12, Fig. 11 (showing teeth with orthodontic wires visible).



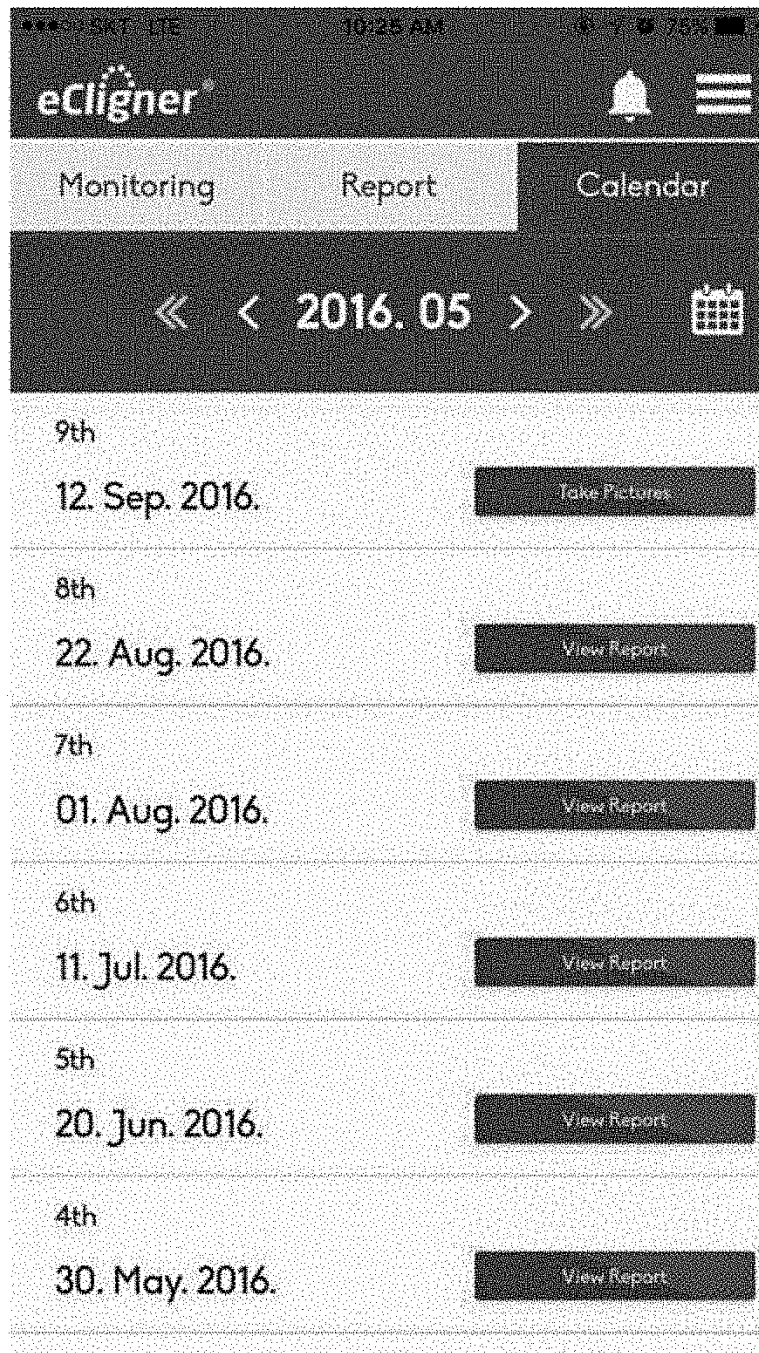
EX1008, Fig. 11. Thus, Salah also teaches acquiring images of a patient's teeth, including teeth of a patient wearing orthodontic appliances.

90. Using a cellphone to acquire an image depicting teeth wearing an aligner is also consistent with published methods of orthodontic practice. For example, a step-by-step guide for remote orthodontic monitoring published by Dr. Scott Frey no later than May 4, 2017, described how to instruct patients to acquire images for use in assessment of orthodontic treatment status, including taking images of "teeth apart and aligners out," "biting down with aligners removed," and "with teeth apart and aligners in." EX1009, 19. This guide confirms that acquiring digital photographs of patient's teeth, including patient-acquired cellphone images of teeth of patients wearing aligners in a service position, was known to be a useful practice for assessment of a patient's orthodontic treatment status.

91. Kim and Invisalign also teach acquiring the image at least one week after the start of treatment with the aligner. Invisalign teaches evaluating “issues that may arise between receipt of Aligners from Align to the completion of treatment.” EX1006, i. Invisalign confirms the commonsense notion that orthodontic treatments with clear aligners take course over many weeks, and it further teaches that treatment is assessed over such periods. For example, when discussing assessing fit by comparing to a 1 mm threshold for separation between teeth and aligners, Invisalign teaches that the “[a]ligners should be fitting well at the end of the 2-week period.” EX1006, 1. Invisalign thus teaches assessing aligner fit (which can be accomplished, for example, by acquiring and analyzing images of teeth and aligners, as taught by Kim) more than one week after starting aligner treatment.

92. Kim agrees, showing that orthodontic images may be taken recurrently based on calendared appointments. Such a schedule is shown in Fig. 15, reproduced below, which shows a calendar with images recurrently scheduled to be taken over the course of months of treatment.

FIG. 15



EX1005, Fig. 15. Kim's description confirms that the progress of teeth alignment is tracked over the dates depicted in Fig. 15, noting that "the orthodontic patient

may view teeth alignment results per date as illustrated in FIG. 15.” EX1008, ¶83.

Thus, both Invisalign and Kim teach assessing a patient’s orthodontic treatment status, including by acquiring cellphone images of a patient’s teeth wearing aligners, more than one week after starting treatment with the aligner.

3. Step 1(b’')

93. Step (b’’) of claim 1 recites as follows:

b’’) analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for at least one tooth attribute of an “analysis tooth zone” representing, at least partially, a tooth on said analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image

94. I have been instructed that this claim term should be interpreted to encompass the use of a device comprising a deep learning neural network trained with a learning base, performing a function of analyzing an image (i.e., the analysis image). I apply this interpretation in my discussion throughout this declaration. Kim, Invisalign, Salah, and Maninis teach this step.

95. Salah teaches that dental images may be analyzed to detect “discriminant information,” which can include contours of objects depicted in the images, including teeth and other intraoral objects, such as gums. EX1008, 34:5-14, 55:21-56:7. A contour map of detected tooth edges showing the resulting outlines of teeth is shown in Fig. 6b, reproduced below.

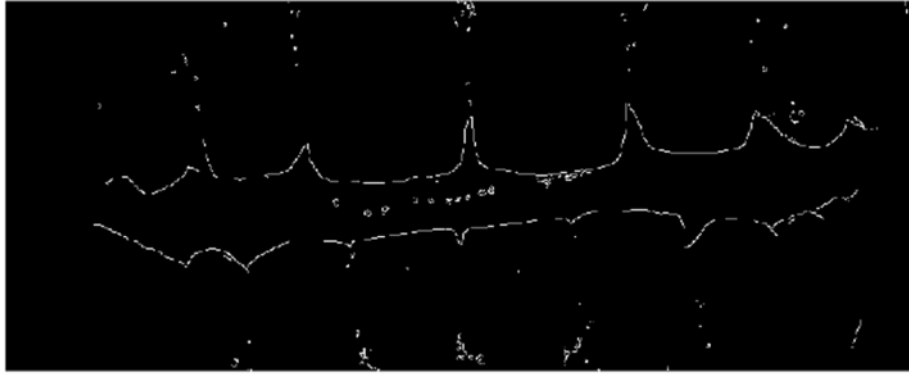


Fig. 6b

EX1008, 34:5-14, Fig. 6b. This discriminant information can be used to generate a “reference map representing the discriminant information,” and Salah provides an example of such a map generated for an image depicting teeth of a patient with an intraoral appliance. EX1008, 40:19-21, 64:7-8; *see id.*, Fig. 11 (reproduced above for element 1(a’)).

96. Salah teaches that its contour detection may use “image processing software” employing any of a variety of known methods, which “can be used, for example, to detect high contrast regions and determine contours.” EX1008, 34:5-36:14. To detect “discriminant information” (e.g., “contour information”), Salah teaches a list of “well-known and preferred methods,” while noting that other methods, though not necessarily preferred, “can also be used,” including analysis “by Deepleraning.” EX1008, 34:5-36:14. A POSA reading Salah would recognize that “Deepleraning” is a typographical error, and that Salah meant to refer to the

use of “deep learning” methods, as the use of deep learning was well-known for various applications including detection of contours. *See generally* EX1007.

97. Maninis provides an example of deep learning techniques a POSA would have found useful for the edge and contour detection taught by Salah. Maninis teaches methods of detecting edges (or “contours”) in images using a deep learning device in order to perform image classification, using a process referred to as “Convolutional Oriented Boundaries (COB).” EX1007, Abstract. Maninis discloses methods of “contour detection” that produce “multiscale oriented contours and region hierarchies” using “Convolutional Neural Networks (CNNs).” *Id.* In particular, Maninis’s deep learning architecture generates a “contour probability map,” which is transformed “into a hierarchical boundary map,” with different probabilities translating to different “strengths” of borders between regions containing objects in the image. EX1007, 3, 5; *see id.*, Figs. 3-5 (depicting contour probability maps used to detect contours between regions).

98. Maninis describes that “latest wave of contour detectors takes advantage of deep learning to obtain state-of-the-art results,” and further notes that “[a]t the core of all these deep learning approaches, lies a *base CNN*,” with Maninis’s COB technique being compatible with different CNNs, applicable in a “modular” fashion that “can incorporate and benefit from future improvements in the base CNN.” EX1007, 2-3 (emphasis original). Maninis describes that its

techniques provide improvements over “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, 2-3. Maninis’s “deep learning architecture” is illustrated in Fig. 2 (below), showing the generation of contours by using multiple layers of feature maps in a convolutional neural network to detect image contents at different scales (fine and coarse detections) as well as different orientations.

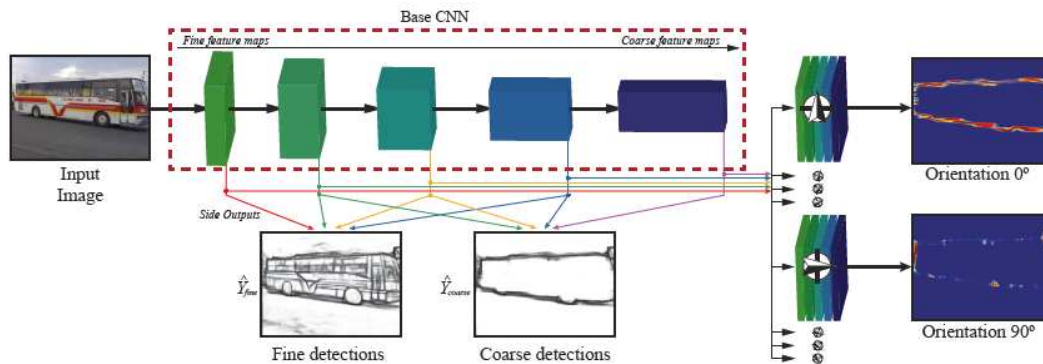


Fig. 2. Our deep learning architecture (best viewed in color). The connections show the different stages that are used to generate the multiscale contours. Orientations further require additional convolutional layers in multiple stages of the network.

EX1007, Fig. 2.

99. Maninis’s use of a computing device comprising a convolutional neural network not only meets the interpretation of step 1(b’’)’s deep learning device that I discuss above, but it more particularly matches the preferred embodiment of a deep learning device for the ’248 patent. EX1007, 3-5, 9. The ’248 patent repeatedly states that a “deep learning device” is “preferably a neural network.” EX1001, 1:24-25, 1:39-40, 1:51-2:3, 3:14-18, 4:5-6, 8:28-40, 13:53-58, 24:3-11, 24:34-35, 24:54-55, 28:17-28, 29:14-15, 29:45-46. Even the specific types

of neural networks overlap between the '248 patent and Maninis: the '248 patent lists “Microsoft ResNet” and “VGG Net” as two examples of neural networks, matching Maninis’s presentation of “results both using VGGNet and ResNet.” EX1001, 16:9-36; EX1007, 2.

100. Maninis describes that, as part of its neural network’s deep learning, the network is trained on a “training dataset” (i.e., a “learning base” as recited in the '248 patent). EX1007, 3. Maninis provides an example using a training dataset of 10,103 images (“the PASCAL Context dataset”) with “localized pixel-wise semantic annotations.” EX1007, 5-6. These annotations are indications, pixel-by-pixel, of the locations of different objects in each image, identifying what object is depicted in each part of the image. To provide “training and optimization,” to the neural network, Maninis divides this set of images into “train, validation, and test sets.” *Id.* These subdivisions of its training dataset (learning base) are then used to train, validate, and test the neural network, demonstrating a substantial improvement over prior classification techniques. *See* EX1007, 5-6, Table 3 (improvement from 24.37 s to 0.79 s), Fig. 6 (showing improved classification accuracy relative to local gradient estimation techniques).

101. 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 (discussing semantic contours and semantic

segmentation). Maninis describes this type of image recognition task as involving “image classification, object detection, and semantic segmentation” where detected contours are assigned to categories corresponding to the “ground-truth” (i.e., the identification of the real-life objects present in both the foreground and background of 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). By labeling sets of pixels with the objects detected to be at the pixels, Maninis produces values (i.e., the values assigned as labels) for the objects in the image.

102. In view of Maninis’s teaching that its techniques for deep-learning-based contour detection can be used to identify and classify objects in an image, a POSA would be motivated to identify teeth and aligners depicted in orthodontic images by detecting their contours and identifying the objects found within each contour. Adjusting Maninis’s deep learning device to perform the specific task of identifying teeth and orthodontic aligners would have been straightforward and well within the skill of the POSA. To accomplish such an application, a POSA would have trained a neural network like Maninis’s using a training dataset representative of the types of images to be categorized (i.e., images of the dental arches of patients wearing aligners in a service position, similar to the images shown in Invisalign, and taken in the context of Kim and Salah’s cellphone

imaging techniques). These images would be labeled with appropriate annotations representing the categories relevant to measurement of objects' locations, such as identifying each tooth, the corresponding parts of the aligner to which distance is to be measured for each tooth, and the part of the image corresponding to the gap between tooth and aligner. A POSA would then have set the output layers of Maninis's neural network to provide the required object identifications corresponding to the labels used to identify individual teeth, aligner portions, and spaces between them, based on Maninis's teaching of semantic boundaries, so that the boundaries of each part of the image could be identified by the neural network after training. Applying the deep-learning based image processing approach taught by Salah and Maninis to detect contours in an image representing a patient's teeth with an aligner worn in a service position on the teeth would result in detection of the edges of the depicted objects (teeth, aligner portions, etc.).

103. To illustrate what I mean by this, below I have provided an example edge detection that I generated based on the image from Invisalign using simpler techniques known prior to Maninis. First I show an image of teeth wearing an aligner in a service position, as shown in Invisalign. Then I have provided a second image showing the result of applying a simple Canny edge detection technique to the Invisalign image, combined with bilateral filtering for noise reduction, consistent with the teachings of Salah. *See* EX1008, 34:11-35:11 (teaching to use

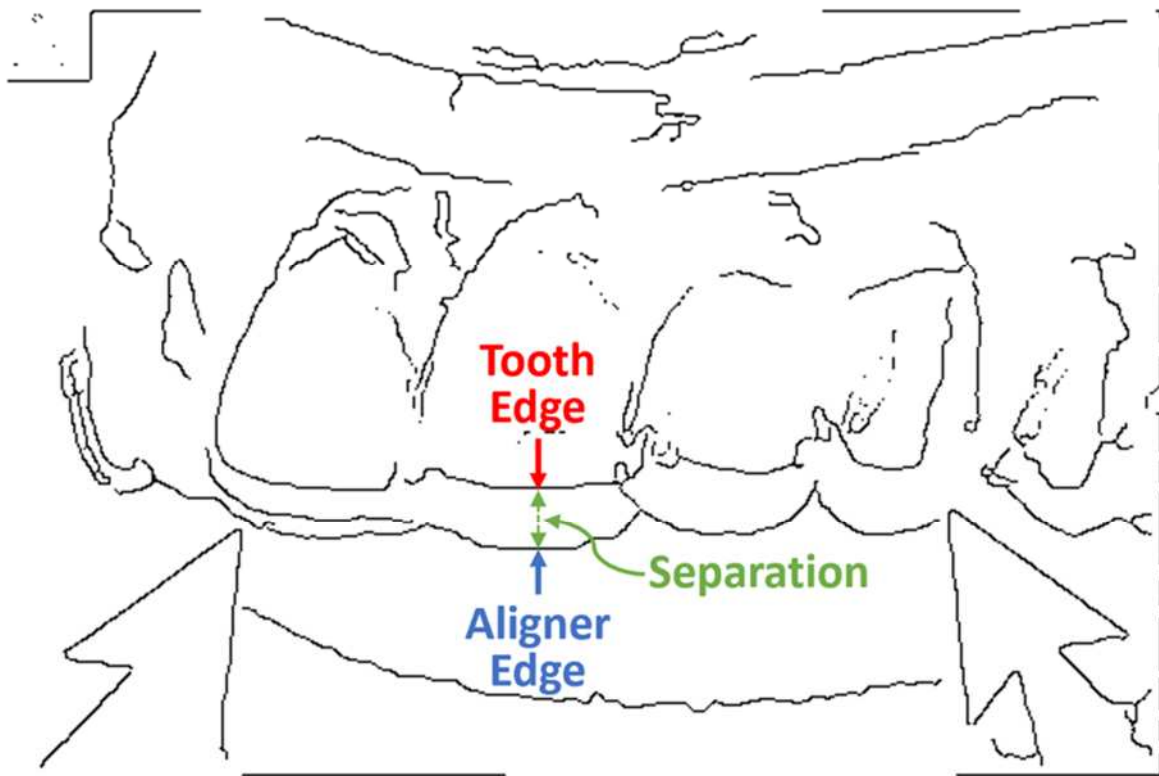
“application of well-known masks or filters provided by image processing software,” including “a bilateral filter” and Canny edge detection). I have also added simple annotations indicating the locations of the edge of a tooth and the edge of a corresponding part of the aligner, as well as showing the separation between the two.

Original Image:



EX1006, 1.

Simple Edge Detection:



To produce the above edge-detected figure, I used a Canny edge detector, which is a much older and less advanced technique than the one provided by Maninis. *See* EX1007, 2, 13 (citing to Canny [24] as an example of an early approach to contour detection that is “unable to handle information captured by richer features such as color or texture”); *see also* EX1008, 34:16 (noting that one may “search for contours using Canny's algorithm”). Canny edge detection produces a simple binary map, in which edge pixels have one value (e.g., 1) and all other pixels have a different value (e.g., 0). In practice, a deep learning contour detector like Maninis’s would provide more accurate and detailed edges, as shown for various objects in Figs. 9 and 10. *See* EX1007, Figs. 9-10; *see also id.*, 7 (improvements

even over state-of-the-art techniques). Going beyond mere detection of edges, Maninis's technique also provides specific identification of the objects found in different regions defined by the detected edges. *See* EX1007, 11-12. As can be understood by extrapolating from my simple example, as well as seen from Maninis's real results, the boundaries detected by Maninis's deep-learning neural network match the contours a person notices naturally when viewing an image. Similarly, the gap that would be detected between the contours in Invisalign corresponds to the gap naturally perceived by a person looking at the image. Since the photographs of Kim, Salah, and Invisalign are ordinary camera images such as those acquired by a cellphone, they show essentially the same thing a person would see when viewing teeth and an aligner in real life. The ability to see a gap and visually assess it in real life naturally maps over to the ability to detect and measure the same gap using computer vision techniques like Salah and Maninis's contour and object detection.

104. As I explained above in section IX.A, a POSA would have been motivated in view of Kim and Invisalign to apply the contour-detection techniques of Salah and Maninis to detect edges of teeth and aligners depicted in images of patient's teeth so that the assessment of gap between aligner and teeth could be made, as taught by Invisalign. Because the detected edges of the aligner and teeth correspond to the edges of the physical objects depicted, a POSA would have

understood that measuring the distance between the outlines of the teeth and the aligner as detected by Salah's image processing software would allow determination of the gap between each tooth and the corresponding part of the aligner.

105. As I noted above, I am interpreting the analysis recited in claim 1(b'') to encompass the use of a device comprising a deep learning neural network trained with a learning base, performing a function of analyzing an image (i.e., the analysis image). Applying this interpretation, the analyzing of the image of a patient's teeth and aligner using the deep-learning contour detection taught by Maninis and Salah would constitute "analysis of the analysis image by means of a deep learning device," as Maninis teaches image analysis with a deep-learning device, in the form of computing device with a trained neural network.

106. Further, using the detected tooth and aligner edges to determine the separation between one or more teeth and the aligner would constitute using the deep-learning device to "determine a value for at least one tooth attribute of an 'analysis tooth zone' representing, at least partially, a tooth on said analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image." Maninis's contour detection detects hierarchies of regions containing objects. EX1007, Abstract, 3-5. Such a region containing a tooth (such as a region containing a tooth

alone or a region containing a tooth and an aligner surface separated by a gap) would constitute an analysis tooth zone at least partially representing a tooth on the analysis image. This is because such a region would literally be a part of an image representing a tooth (at least partially), with the detected boundaries defining the edges of the region.

107. Further, performing this analysis would generate tooth attributes for the tooth zone. These would include, for example, the locations of the surfaces of the tooth and aligner, which would correspond to the detected contours. Additionally, the measured separation between the contours would also be a tooth attribute for the tooth zone (it would be an attribute indicating how far the tooth in that zone is from the aligner). *Id.* Each of these tooth attributes would be a tooth attribute related to the separation between the represented tooth and aligner. For example, the location of a tooth contour relates to the separation (in that it is from that contour that a gap would be measured), and the location of the corresponding aligner contour would relate to the separation in the same way. And of course, the measure of separation between contours (consistent with the teachings of Invisalign) is directly related to the separation, since it is a measure of that separation.

108. In sum, Kim, Invisalign, Salah, and Maninis teach performing analysis of the analysis image to determine tooth attribute values for the edges of

teeth and aligners, to determine value for a separation between the two, and these functions would be performed by means of a deep learning device trained by means of a learning base as taught by Salah and Maninis. Accordingly, Kim, Invisalign, Salah, and Maninis teach step 1(b").

109. Accordingly, it is my opinion that claim 1 would have been obvious over Kim, Invisalign, Salah, and Maninis.

C. Claim 2

110. Claim 2 recites "The method as claimed in claim 1, comprising, after the step b"), step c") as follows: c") assessment of the suitability of the aligner as a function of the value of said tooth attribute." Kim, Invisalign, Salah, and Maninis teach this step.

111. As I discussed above for claim 1, Invisalign teaches to use a threshold to assess aligner acceptability. For example, Invisalign teaches that when an aligner worn on a patient's teeth has a tooth-aligner gap of less than 1 mm, "this is natural deflection that should occur so the Aligner can move the tooth," whereas if the gap is greater than 1 mm, this means that the "Aligners are not seating well." EX1006, 1. In view of this teaching, a POSA would have been motivated to compare the values discussed above for claim 1 (that is, the values representing the surfaces of the patient's teeth, the corresponding aligner surfaces, and the separation between them) to apply the acceptability threshold taught by Invisalign,

thereby determining that where the threshold is exceeded, the aligner is not seated well on the teeth, or conversely that where the threshold is not exceeded, the aligner is acceptably seated.

112. This assessment would be consistent with Kim’s teaching to “check[] a photo or a video of the clear aligner currently worn by the orthodontic patient, and determine[] whether the clear aligner is normal and corresponds to an appropriate step.” EX1005, ¶66. For example, if the result of applying Invisalign’s threshold assessment indicated that one or more teeth had excessive gaps, that would be an indication that the aligner was not normal or did not correspond to an appropriate step. EX1006, 1 (evaluation may indicate aligner has “excessive flash” or results from a “distorted impression” (meaning it was not normal)), 6 (lack of fit may indicate incorrect aligners or need for mid-course correction (meaning the aligner did not correspond to an appropriate step)).

113. A POSA would have understood that the assessment could be implemented in multiple ways. As a simple example, the edge detection software could indicate the locations of the aligner and tooth edges, or the gap between (e.g., measured in pixels), and provide that information to the orthodontist to allow the orthodontist to make the judgment of whether the aligner was suitable. As another option, the threshold taught by Invisalign would be applied computationally, by comparing the measured aligner-tooth separation to the

threshold. A POSA would have known that a conversion of a length threshold like 1 mm to a distance measurement in an image (e.g., in pixels) would be straightforward. For example, a known width (like the typical size of a tooth, or the known width of the patient's tooth, based on a model as taught by Salah) would permit a conversion of units using simple geometry. Further, Salah teaches that registration marks with known size and spacing can be included in the images, and these too would provide a straightforward calibration basis to convert image measurements in pixels to metric measurements in units of distance (e.g. millimeters). *See* EX1008, 38:7-9.

114. Accordingly, it is my opinion that claim 2 would have been obvious over Kim, Invisalign, Salah, and Maninis.

D. Claim 4

115. Claim 4 depends from claim 2, reciting “The method as claimed in claim 2, comprising, after the step c”), a step d”) as follows: d”) sending of an information message as a function of said assessment.” Claim 4 would have been obvious in view of Kim, Invisalign, Salah, and Maninis.

116. A POSA would have found it obvious to send an information message as a function of the assessment recited in claim 2 (i.e., an assessment of the aligner's suitability). For example, it would have been obvious to send a message indicating that the aligner was or was not suitable based on the assessment,

because otherwise the patient (or treating orthodontist) would not be aware of the assessment. As a practical matter, it would have been common sense for the results of any assessment performed by a computer, such as the assessment discussed above for claim 2, to be sent as an information message, since the purpose of performing evaluations of aligner fit is to inform the people involved in patient treatment, such as the orthodontist and the patient, about the results of the analysis.

117. As one example, the assessment could be used to provide guidance to a patient about wearing the aligner. Invisalign teaches that poor fit (indicated by a tooth-aligner separation greater than 1 mm) may be due to “non-compliance,” i.e., the “patient not wearing Aligners long enough for [a] tooth to move.” EX1006, 1. In this case, the guide teaches that the patient should be told to extend wear time of the aligner, and that orthodontists should emphasize to the patient “never to dispose of their Aligners and instruct them that the Aligners should be fitting well at the end of the 2-week period.” *Id.* Sending this instruction would constitute sending an information message as a function of the assessment.

118. For example, this message could be sent to the patient’s mobile device. This option is taught by Kim, which teaches that a “teeth alignment guidance information transmission program” can send information from the orthodontist to the patient’s mobile device that “advises, for example, an appropriate step of a clear aligner to be currently worn by the orthodontic patient.”

EX1005, ¶63; *see also id.*, ¶64 (“[T]he teeth alignment guidance information may include at least one of a clear aligner information about a clear aligner to be worn in a current step among a clear aligner set, a wearing time information about a wearing time of the clear aligner, a replacement information about replacement of a part of or the whole clear aligner set, a clear aligner state information about the state of the clear aligner, a clear aligner usage guidance information....”). Sending a message to the patient in this manner, such as a message to extend aligner wear time, to not dispose of them, and that they should fit well within two weeks, would constitute sending the information message as a function of the assessment as recited in claim 4.

119. Accordingly, it is my opinion that claim 4 would have been obvious over Kim, Invisalign, Salah, and Maninis.

E. Claim 7

120. Claim 7 depends from claim 1 and further recites that step (b”) comprises four additional steps. As I discuss below, Kim, Invisalign, Salah, and Maninis teach each of these steps.

1. Step 7(1)

121. Step 1 of claim 7 recites “creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image representing an aligner worn by a “historical” patient and comprising one or more

zones each representing a tooth, or “historical tooth zones”, to each of which, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, a tooth attribute value is assigned.” Kim, Invisalign, Salah, and Maninis teach this element.

122. To begin with, a POSA would have understood that a large number of images would be needed to train any deep learning device. Indeed, a POSA would have recognized that far more than 1000 images would typically be contained in a learning base (i.e., a training dataset) from which images were used to train such a device. For example, Maninis discloses training its neural network on images selected from a training dataset (corresponding to the ’248 patent’s “learning base”) of over 10,000 images. *See* EX1007, 1 (using “a trainval set with more than 10 000 challenging and varied images”), 5-6 (“We extend the main BSDS benchmarks to the PASCAL Context dataset, which contains carefully localized pixel-wise semantic annotations for the entire image on the PASCAL VOC 2010 detection trainval set. This results in 459 semantic categories across 10,103 images, which is an order of magnitude (20×) larger than the BSDS.”). Thus, as specifically taught by Maninis, a POSA would have considered it natural to train a deep learning device with more than 1000 images.

123. With regard to the images being “images of dental arches, or ‘historical images,’” a POSA would have understood that this would be a routine

manner of adapting Maninis's contour and object detection device to the context of Kim, Invisalign, and Salah. Maninis discloses using a training set with 459 semantic categories corresponding to the objects depicted in the images. EX1007, 5-6. The semantic category depends on the object depicted in each image; for example, the images shown in Fig. 9 depict images containing animals (a dog, a cat, etc.), foreground objects (a dog toy, an airplane, etc.), and background objects (a fence, trees, the sky, etc.). *See* EX1007, 9, Fig. 10; *see also id.*, Tables 4-7 (listing several object categories used for identifying particular objects in images). Training a neural network on images of such objects lets the network learn to detect the categories of objects it has seen during training. Thus, when training a neural network to perform image recognition and edge detection on images of patient's teeth wearing aligners, a POSA would have understood it to be important to include images of patients' teeth wearing aligners in the training set. Indeed, to most efficiently train such a neural network, the training set should include images representative of the types of images likely to be seen during ordinary use of the imaging device, such as real historical photographs taken of patient's teeth wearing aligners with the device in which the trained network will be deployed (e.g., a cellphone as taught by Kim and Salah).

124. Regarding the requirement that "each historical image compris[e] 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,” a POSA would have performed this as well based on the teachings of Kim, Invisalign, Salah, and Maninis. As discussed above for claim 1, a POSA would have been motivated to identify teeth, aligners, and the gaps between them in the image by using Maninis’s object detection algorithms to assign tooth attributes to imaged teeth. In order to train a neural network to detect such attributes, the training set would be labeled with the attributes exhibited by objects depicted in each training image. Maninis teaches this practice, noting that its training set images contain “carefully localized pixel-wise semantic annotations for the entire image,” which “results in 459 semantic categories” being assigned to objects depicted within the set of 10,103 images. EX1007, 5-6. Maninis’s “pixel-wise semantic annotations” are indications of which pixels correspond to each object in a given image, with the shape of a given object indicated by a change from one annotation to another at the boundary between objects. Applying this same concept to a training set appropriate to the dental imaging application of Kim, Invisalign, and Salah, a POSA would have applied annotations to historical dental images to indicate the attributes of each object in the image, including assigning tooth attributes to each tooth in each image to each tooth in each image, as well as to other parts of the image including the aligner and the gap separating the aligner and the tooth.

125. With labels to indicate the pixels corresponding to teeth, the pixels corresponding to aligners, and the pixels corresponding to the gap between the two, each historical image would include tooth attribute values assigned directly to the image by the labels. For example, the boundaries between regions would directly indicate the edges of teeth, or of the aligner, both of which are tooth attributes relating to the separation between the tooth and the aligner (as I explained above, these edges relate to the separation at least because they define the edges between which the separation is measured). Further, labeling the pixels between each tooth and the corresponding aligner surface (i.e., the gap), the image would include a direct label for the tooth of the separation between the tooth and the aligner—the labeled gap is literally a label of the separation to be detected in order to apply Invisalign’s teaching to assess the separation between teeth and the corresponding parts of the aligner.

126. Thus, Kim, Invisalign, Salah, and Maninis teach each part of step 1 of claim 7.

2. Step 7(2)

127. Step 2 of claim 7 recites “training of at least one deep learning device by means of the learning base.” Kim, Invisalign, Salah, and Maninis teach this element.

128. As discussed above for claim 1, Maninis teaches that its neural network is trained using a training dataset, which satisfies the meaning of “learning base” as used in the ’248 patent. *See, e.g.*, EX1007, 3-4, 7-8. In the context of Kim, Invisalign, Salah, and Maninis, the learning base discussed above for step 7(1) containing annotated historical images of dental arches would be used to train a neural network like the one disclosed in Maninis. Thus, Kim, Invisalign, Salah, and Maninis teach training “at least one deep learning device by means of a learning base,” as recited in step 7(2).

3. Step 7(3)

129. Step 3 of claim 7 recites “submission of the analysis image to the deep learning device for it to determine at least one probability relating to: the presence, in a location of said analysis image, of an analysis tooth zone, and the attribute value of the tooth represented on said analysis tooth zone.” Kim, Invisalign, Salah, and Maninis teach this element.

130. As I discussed above for claim 1, it would have been obvious to submit images of patients’ dental arches, as captured by the cellphone of Kim and Salah, to be analyzed by a deep learning neural network as taught by Maninis to detect regions containing teeth, aligners, and the gap between, and to characterize their attributes, including the separation between teeth and the corresponding parts

of the aligner, to accomplish the aligner fit evaluation taught by Invisalign.

Maninis teaches the determination of probabilities related to these properties.

131. Maninis teaches the determination of probabilities related to “the presence at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone.’” For example, Maninis teaches that its contour detection procedure is built “on the concept of Ultrametric Contour Map (UCM), which transforms a contour detection probability map into a hierarchical boundary map, which gets partitions at different granularities when thresholding at various contour strength values.” EX1007, 5 (citation omitted). This framework involves producing “scale-specific contour images (side outputs) for different scales of a network, and combines their activations linearly to produce a contour probability map.” EX1007, 3. In essence, Maninis is describing a process that assigns a probability to locations within an image, with the assigned probabilities indicating the probability that a contour is located along a given set of pixels (a “contour probability map”). Maninis goes on to provide a set of equations by which the “probability $P(\cdot)$ is obtained.” EX1007, 3. These probability maps are then used to derive the contours most likely to be present in the image, using techniques adapted from prior techniques to transform “a contour detection probability map into a hierarchical boundary map.” EX1007, 5. Maninis describes the different probabilities as being treated as different “strengths” of boundaries,

depicted in the figures as red lines of different thickness. *See* EX1007, 5, Figs. 3, 4-5 (depicting a simple portion of an image with three boundaries, having strengths of 0.6, 0.2, and 0.6). By comparing these “strengths” derived from the contour probability map, Maninis determines a “hierarchy of regions” dividing the different objects in an image. *Id.*

132. In the context of Kim, Invisalign, and Salah, this contour detection process involves determining a probability related to “the presence at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone,’” because the set of tooth edges define analysis tooth zone that has been determined to contain a particular tooth. Since Maninis’s neural network generates probabilities in the process of defining those edges, it generates probabilities related to the presence of the zone representing the tooth.

133. With regard to “the attribute value of the tooth represented on said analysis tooth zone,” as discussed above for claim 1 (as well as the other steps of claim 7), in the combined system of Kim, Invisalign, and Salah, the detected teeth of a patient in an image would be assigned tooth attributes such as tooth and aligner edges and the size of the gap between, as represented visually in the image and determined by applying a contour and object detection process as taught by Maninis.

134. Thus, Kim, Invisalign, Salah, and Maninis teach each part of step 3 of claim 7.

4. Step 7(4)

135. Step 4 of claim 3 recites “determination, as a function of said probability, of the presence of a separation of the aligner and of the tooth represented by said analysis tooth zone.” Kim, Invisalign, Salah, and Maninis teach this element.

136. Maninis detects regions based on probabilities, as I discussed above. Further, as discussed above for claim 1, the edges of the regions correspond to the edges of teeth and aligners, and the distance spanned by the gap between these edges corresponds to the separation between the tooth and the corresponding part of the aligner, which is what Invisalign teaches to measure to assess aligner fit. EX1006, 1. Determining the size of a separation would include determining the presence of a separation, as if a separation of finite size is detected, then the separation is present, whereas if a separation is so small that it is not detectable (e.g., if no gap was present, such that the separation was zero), then the presence of a separation would not be detected. For example, in an image like that shown in Invisalign to demonstrate poor fit as exemplified by a gap greater than 1 mm, the presence of a separation would be detected, as can be understood intuitively by the fact that you can see a gap easily by eye. Thus, applying the teachings of Kim,

Invisalign, Salah, and Maninis would involve detecting the locations and edges of teeth and aligners using probabilities, and then determining based on those probabilities the presence of a separation of the aligner and the tooth. Accordingly, Kim, Invisalign, Salah, and Maninis teach step 4 of claim 7.

137. Accordingly, it is my opinion that claim 7 would have been obvious over Kim, Invisalign, Salah, and Maninis.

F. Claim 8

138. Claim 8 depends from claim 1 comprises three additional steps. These steps resemble the steps recited in claim 7, but, among other differences, the steps recite an “image attribute,” rather than the tooth attributes recited in claims 1 and 7. As discussed above in section VIII.B, the image attribute recited in claim 8 encompasses any attribute of the image as a whole that relates to the separation between one or more teeth and the aligner. This attribute may be a single value or a set of values, including a value or values derived from tooth attribute values. Kim, Invisalign, Salah, and Maninis teach each of the steps of claim 8, as I discuss below.

1. Step 8(1')

139. Step 8(1') is the same as step 7(1), except the historical images must comprise an image attribute value rather than historical tooth zones and tooth attributes. As I explain above for step 7(1), a POSA would have been motivated to

create a learning base of more than 1000 historical images of dental arches of patients wearing aligners to train a neural network. *Supra* section IX.E.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 Kim, Invisalign, and Salah, these images would be images representative of the type acquired with Kim and Salah's cellphones in view of the teachings of Invisalign. In other words, these would be images of patient's teeth with aligners in service positions, captured by a cellphone. For the same reasons as I described above for claims 1 and 7, using these types of "representative" images would be expected to efficiently train a neural network to recognize depicted objects and their contours. *See supra* sections IX.A-B, IX.E. Thus, for example, each acquired historical image would have an aligner and teeth represented in the image, with annotations indicating the regions containing each tooth and corresponding aligner portion, as well as indicating the separation between each tooth and corresponding aligner part (i.e., the gap between the two).

140. The values for these attributes, as reflected directly in the annotations themselves, would relate to the separation of the teeth and the aligner. For example, the '248 patent indicates that the "presence" of an aligner can be an image attribute value, and that an image attribute value can also relate to the

presence of a separation. *See* EX1001, 26:11-14, 29:30-53. An annotation of an aligner relates to the presence of the aligner; likewise, an annotation of a gap between an aligner and a tooth relates to the presence of a separation, so images annotated as containing an aligner and/or a gap would be annotated with image attributes indicating that an aligner and/or gap was present. As a further example, the the locations of each tooth and each corresponding aligner part would also be image attributes, as they relate the tooth-aligner separation, since that separation is a measure of the distance from the edge of the region containing the tooth to the edge of the region containing the aligner. Considered collectively, the set of annotations would relate to an image attribute value, as they would relate to the value of whether the aligner in the image was well or poorly fitted, under the standard taught by Invisalign (e.g., they would depict gaps greater or less than 1 mm). The gap between aligner and tooth directly represents the separation between each tooth and the aligner, so the annotation of such gaps would also indicate an attribute value for the image relating to the separation between “the tooth or teeth and the aligner represented on the historical image.”

141. Thus, Kim, Invisalign, Salah, and Maninis teach step (1') of claim 8.

2. Step 8(2')

142. Step 8(2') recites “training of at least one deep learning device, by means of the learning base.” This is identical to the language recited in step 7(2).

Accordingly, this step would have been obvious for the reasons I discussed above for step 7(2). *See supra* section IX.E.2.

3. Step 8(3')

143. Step 8(3') recites “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.” This step resembles step 7(3), except that the probability relates to an image attribute value rather than the presence of an analysis tooth zone and corresponding tooth attribute.

144. As I discussed above for step 7(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* section IX.E.3. The probabilities relate to the image attributes discussed above for step 8(1'), such as the distance between the teeth's edges and the corresponding edges of the aligner, as reflected in the boundaries of the respective tooth and aligner regions as well as in the annotation of the gap between teeth and aligner. For example, the probabilities used to determine the boundaries defining the ends of each tooth and the corresponding surfaces of the aligner relate to various image attribute values that themselves relate to separation between aligner and teeth. Most obviously, they relate to the separations of all teeth from the aligner (and as I discuss above in section VIII.B, an image attribute can be a list of tooth attributes), but they also

relate to properties used to assess those separations, such as a list of locations of all teeth edges and a list of locations of all aligner edges. Additionally, the detection of the presence of aligners and the gap between the aligners and teeth relate would be image attributes (as discussed for step 8(1')) that are related to the separation. *See* EX1007, 11-12 (identifying objects present in regions of images). I further discuss various attributes of these types below for claim 15 in section IX.I, when I discuss various image attributes that a POSA would also have had reason to detect.

145. Accordingly, it is my opinion that claim 8 would have been obvious over Kim, Invisalign, Salah, and Maninis.

G. Claim 13

146. Claim 13 depends from claim 1 and further recites that “the deep learning device is a neural network.”

147. As discussed above for claim 1, Maninis teaches that its image analysis for contour and object detection is performed using a convolutional neural network (CNN). EX1007, Abstract. Indeed, Maninis teaches that the use of a CNN is the core of “deep learning approaches” like its own, and it depicts using a single-pass CNN as the first element of its analysis procedure. EX1007, Abstract, Fig. 1 (reproduced below); *see also id.*, 1-2, 13, Fig. 2.

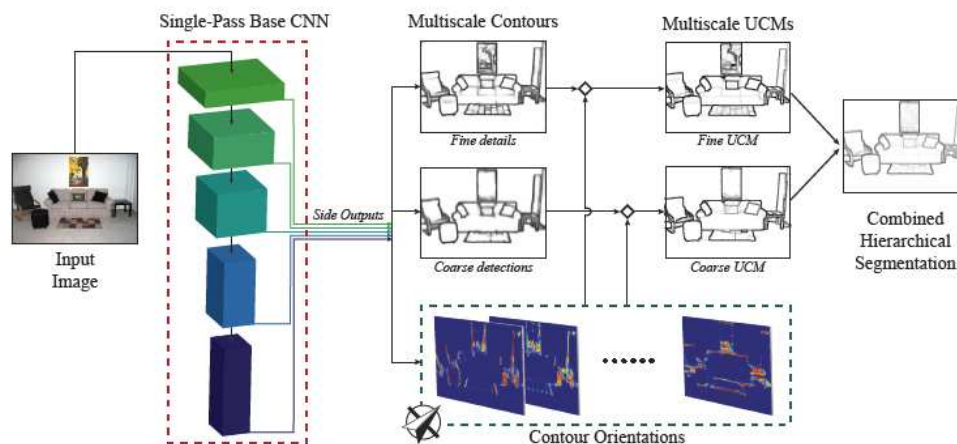


Fig. 1. **Overview of COB:** From a single pass of a base CNN, we obtain multiscale oriented contours. We combine them to build Ultrametric Contour Maps (UCMs) at different scales and fuse them into a single hierarchical segmentation structure.

EX1007, Fig. 1.

148. Accordingly, it is my opinion that claim 13 would have been obvious over Kim, Invisalign, Salah, and Maninis.

H. Claim 14

149. Claim 14 depends from claim 1 and further recites that step (b'') comprises four additional steps. These steps very closely resemble the steps of claim 7. In fact, steps 14(1), 14(2), and 14(3) are identical to steps 7(1), 7(2), and 7(3), respectively. Thus, Kim, Invisalign, Salah, and Maninis teach each of these steps for the reasons I discussed above for claim 7. *See supra* sections IX.E.1-3.

150. With regard to step 14(4), it resembles step 7(4), but differs slightly in wording. Claim 7(4) recites “determination, as a function of said probability, of the presence of a separation of the aligner and of the tooth represented by said analysis tooth zone.” Step 14(4) recites “determination, as a function of said probability, of an amplitude of said separation.” Thus, whereas step 7(4) recites determining the

presence of a separation, step 14(4) recites determining an *amplitude* of the separation. As I discuss below, Kim, Invisalign, Salah, and Maninis teach step 14(4) for reasons similar to those I discussed above for step 7(4). *See supra* section IX.E.4.

151. As I discussed above in section IX.E, Maninis teaches detecting objects (which in the context of Kim, Invisalign, and Salah would include teeth, aligners, and spaces between teeth and aligners). *See* EX1007, 11-12. Invisalign teaches assessing aligner fit by comparing the size of separation between teeth and the corresponding part of the aligner to a threshold, where, for example, a separation of greater than 1 mm would indicate a poor fit. EX1006, 1. As I discussed above for claim 1, a POSA would have been motivated by Invisalign to assess the aligner-tooth separation for the patient as reflected in images of the patient's teeth and aligner, so as to determine whether the patient's aligners fit well. *See supra* section IX.A-B. This determination would involve determining an amplitude of the separation between teeth and aligners, as a function of the probabilities recited in step 14(3), as Maninis teaches the generation of probabilities as part of a determination of the boundaries between regions corresponding to different objects in an image. *See supra* section IX.E.3 (discussing identical step 7(3)); EX1007, 3-5, Figs. 3-5. Since Maninis teaches using probabilities to determine edges, and the distance used by Invisalign to

assess aligner fit corresponds to a distance between those edges, the amplitude of that separation would be determined as a function of the probability. Thus, Kim, Invisalign, Salah, and Maninis teach step 14(4).

152. Accordingly, it is my opinion that claim 14 would have been obvious over Kim, Invisalign, Salah, and Maninis.

I. Claim 15

153. Claim 15 is an independent claim reciting similar limitations to claim

1. The preamble and step (a'') of claim 15 recite as follows:

A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a'') more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image" the analysis image being a photograph, or an image extracted from a film;

154. These elements are word-for-word identical to the preamble and step (a'') of claim 1. Claim 15 differs from claim 1 for these elements only in that whereas claim 1 requires that step (a'') be performed with a cellphone, claim 15 does not require that a cellphone be used. The omission of the requirement of a cellphone, however, merely means that the use of a cellphone is optional in claim 15. Thus, my analysis above in sections IX.A-B.2 regarding motivation, expectation of success, and the preamble and step (a'') of claim 1 applies equally to

the preamble and step (a'') of claim 15. These elements are thus taught by the prior art for the same reasons.

155. Step (b'') of claim 15 recites as follows:

b'') analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for an image attribute of the analysis image, the image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on said analysis image.

This step is the same as step 1(b''), except that instead of determining a value for a tooth attribute of a tooth in an analysis tooth zone, claim 15 requires that the analysis using a deep learning device must be used when determining a “value for an image attribute of the analysis image, the image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on said analysis image.”

156. As I discussed above for claims 1 and 8, it would have been obvious to train a neural network, as taught by Maninis and Salah, to identify and detect edges of teeth and aligners, in order to assess the gap between the teeth and aligner, as taught by Invisalign, in order to determine whether the clear aligner is normal and corresponds to an appropriate step, as taught by Kim. *See supra*, sections IX.A, IX.B.3, IX.F. This would include detecting image attributes for the image relating to the separation between the aligner and the represented teeth. For

example, the detected image attributes would include the presence of an aligner in the image and the presence of a gap between that aligner and the teeth. Further detected image attributes would include, for example, the locations of the occlusal edges of each of a plurality of teeth depicted in the image that define the gap. A POSA would also have understood that the detection of gaps between teeth and aligner would allow detection of a more general image attribute value, such as a value characterizing the overall size of the gap (e.g., for each tooth, as an average across teeth, or as a largest tooth gap), any of which would permit generating a value indicating whether the gap was larger or smaller than the 1 mm threshold taught by Invisalign.

157. Detecting such image attributes, based on the deep learning analysis taught by Salah and Maninis, would have been well within the skill of the POSA and would represent no more than commonsense, routine ways of characterizing gaps between aligners and teeth. In essence, each image attribute would be a simple measurement of distance, either individually or collectively, for the set of depicted teeth and the aligner, and each attribute would serve to inform Kim's assessment of whether the aligner was normal and corresponded to an appropriate step, by applying Invisalign's teaching to compare measured distances to a threshold indicating good or poor fit. Each attribute would thus relate "to a separation between at least one tooth represented on the analysis image, and the

aligner represented on said analysis image,” as recited in step 15(b”). Additionally, each attribute would be determined from the detected contours, which in turn would be determined using a deep learning device trained by a learning base, as taught by Maninis. Accordingly, Kim, Invisalign, Salah, and Maninis teach step 15(b”).

158. Accordingly, it is my opinion that claim 15 would have been obvious over Kim, Invisalign, Salah, and Maninis.

J. Claim 16

159. Claim 16 depends from claim 15, reciting “The method as claimed in claim 15, comprising, after the step b”), step c”) as follows: c”) assessment of the suitability of the aligner as a function of the value of said image attribute.” The combination of Kim, Invisalign, Salah, and Maninis teaches claim 16.

160. Claim 16 resembles claim 2, in that just as claim 2 recites assessing suitability of the aligner as a function of the tooth attribute of claim 1, so does claim 16 recite assessing suitability of the aligner as a function of the image attribute of claim 15. Thus, for similar reasons as I discussed above for claim 2, it also would have been obvious assess suitability of an aligner as a function of an image attribute, such as those discussed above for claim 15. *See supra* section IX.C.

161. For example, Invisalign teaches that when an aligner worn on a patient's teeth has a tooth-aligner gap of less than 1 mm, "this is natural deflection that should occur so the Aligner can move the tooth," whereas if the gap is greater than 1 mm, this means that the "Aligners are not seating well." EX1006, 1. In view of this teaching, a POSA would have been motivated to compare the values discussed above for claim 15 (that is, values representing the surfaces of the patient's teeth, the corresponding aligner surfaces, and the separation between them, such as values of each tooth's separation, the largest separation, an average separation, or a number of teeth exceeding 1 mm of separation) to apply the acceptability threshold taught by Invisalign, thereby determining that where the threshold is exceeded, the aligner is not seated well on the teeth, or conversely that where the threshold is not exceeded, the aligner is acceptably seated.

162. Addressing various examples I listed above for claim 15, assessment of suitability can be assessed as a function of the presence of the aligner and of a gap between these the aligner and the teeth, the width of the gap, and the distances of each tooth from its corresponding part of the aligner, as each involves an image attribute relating directly or indirectly to separation. For example, no aligner means a measure is impossible, no gap indicates the aligner meets Invisalign's standard for suitability, and the measures of gap width are directly applicable to that standard (as they may be below or above the threshold taught by Invisalign). The

other properties I discussed above for claim 15 would similarly aid in making the assessment taught by Invisalign, as they indicate, for example, maximal or average gap size, or simply a list of all gaps' sizes. These could be used directly by an orthodontist or by a program aiding the orthodontist to assess aligner suitability.

163. This assessment would be consistent with Kim's teaching to "check[] a photo or a video of the clear aligner currently worn by the orthodontic patient, and determine[] whether the clear aligner is normal and corresponds to an appropriate step." EX1005, ¶66. For example, where applying Invisalign's threshold assessment indicated that one or more teeth had excessive gaps, either individually or collectively, that would be an indication that the aligner was not normal or did not correspond to an appropriate step. For example, Invisalign teaches that the evaluation of separation may indicate aligner has "excessive flash" or results from a "distorted impression" (meaning it was not normal). EX1006, 1. In another example, Invisalign teaches that lack of fit may indicate that incorrect aligners were provided or that the patient may need mid-course correction (meaning the aligner did not correspond to an appropriate step). EX1006, 6. Each such assessment is an assessment, as a function of the image attribute recited in claim 15, of the suitability of the aligner.

164. Accordingly, it is my opinion that claim 16 would have been obvious over Kim, Invisalign, Salah, and Maninis.

165. For the reasons I have provided above, it is my opinion that each of claims 1-2, 4, 7-8, and 13-16 of the '248 patent would have been obvious over Kim, Invisalign, Salah, and Maninis.

X. CONCLUDING STATEMENTS

166. In signing this declaration, I understand that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

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

Dated: October 16, 2023

By: /Hassan Foroosh, Ph.D./
Hassan Foroosh, Ph.D.

XI. APPENDIX – LISTING OF CLAIMS

1. A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a") more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image", the analysis image being a photograph, or an image extracted from a film;

b") analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value

for at least one tooth attribute of an "analysis tooth zone" representing, at least partially, a tooth on said analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image,

in the step a"), a cellphone is used to acquire the analysis image.

2. The method as claimed in claim 1, comprising, after the step b"), step c") as follows:

c") assessment of the suitability of the aligner as a function of the value of said tooth attribute.

4. The method as claimed in claim 2, comprising, after the step c"), a step d") as follows:

d") sending of an information message as a function of said assessment.

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 representing an aligner worn by a "historical" patient and comprising one or more zones each representing a tooth, or "historical tooth zones", to each of which, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, a tooth attribute value is assigned;

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: iii) for each of said teeth, determination of a distance score dependent on

the presence, in a location of said analysis image, of an analysis tooth zone, and

the attribute value of the tooth represented on said analysis tooth zone;

4) determination, as a function of said probability, of the presence of a separation of the aligner and of the tooth represented by said analysis tooth zone.

8. 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", the image attribute relating to a separation between the tooth or teeth and the aligner represented on the historical image;

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.

13. The method as claimed in claim 1, in which in step b''), the deep learning device is a neural network.

14. 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 representing an aligner worn by a "historical" patient and comprising one or more zones each representing a tooth, or "historical tooth zones", to each of which, for at least one tooth attribute relating to a separation between the tooth represented by

the historical tooth zone considered, and the aligner represented, a tooth attribute value is assigned;

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, in a location of said analysis image, of an analysis tooth zone; and

the attribute value of the tooth represented on said analysis tooth zone;

4) determination, as a function of said probability, of an amplitude of said separation.

15. A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a") more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image" the analysis image being a photograph, or an image extracted from a film;

b") analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value for an image attribute of the analysis image, the image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on said analysis image.

16. The method as claimed in claim 15, comprising, after the step b"), step c") as follows:

c") assessment of the suitability of the aligner as a function of the value of said image attribute.

XII. APPENDIX – EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 11,049,248 to Salah et al.
1003	Curriculum Vitae of Hassan Foroosh, Ph.D.
1005	U.S. Patent Pub. No. 2017/0281313 to Taewon Kim
1006	Align Technology, Invisalign Clinical Monitoring Guide, (archived September 26, 2003 at http://web.archive.org/web/20030926024959/http://www.invisaligncec.com:80/consistent/pdfs/ICMG_Final.pdf)
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	Scott Frey, Blog – The Ortho Cosmos, (archived May 4, 2017 at http://web.archive.org/web/20170504155213/http://theorthocosmos.com:80/blog)
1012	PCT Pub. No. WO 2016066651 to Salah et al.
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)
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)