

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

REPLY DECLARATION OF DR. HASSAN FOROOSH

TABLE OF CONTENTS

I. QUALIFICATIONS 1

II. LEGAL STANDARDS 1

III. LEVEL OF SKILL IN THE ART 2

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

 A. The Combination of Kim, Invisalign, Salah, and Maninis is Not
 Inoperable 4

 B. A POSA Would Have Had Reason to Use Salah and Maninis’s
 Deep Learning Techniques to Detect Distances Between Teeth
 and Corresponding Aligner Edges with a Reasonable
 Expectation of Success 7

 C. A POSA Would Not Have Been Deterred by Development
 Costs for the Proposed Combination 14

V. CONCLUDING STATEMENTS 20

VI. APPENDIX – CITED PETITIONER EXHIBITS 21

VII. APPENDIX – CITED PATENT OWNER EXHIBITS 23

I, Hassan Foroosh, declare as follows:

I. QUALIFICATIONS

1. I submitted an expert declaration in support of Petitioner Align's petition for *inter partes* review (IPR) of U.S. Patent No. 11,049,248 (the "'248 Patent," EX1001) in Case No. IPR2024-00052. I understand that an IPR was instituted and Patent Owner Dental Monitoring submitted a Patent Owner Response. I submit this expert declaration in support of Align's Reply to the Patent Owner Response.

2. I previously set forth my qualifications in my first declaration (EX1002) at paragraphs 1-13. I also submitted my curriculum vitae (EX1003) at that time, which further sets forth my qualifications.

3. In preparation for this declaration, I have reviewed the Declarations of Dr. John Mongan (EX2002 and EX2028) and the Declaration of Dr. Budi Kusnoto (EX2030). Lists of exhibits I cited in this declaration are provided in the Appendices.

II. LEGAL STANDARDS

4. I understand that my opinions in this declaration require applying the legal principles outlined in my first declaration. As was done previously, I have applied those same legal principles in forming my opinions here.

III. LEVEL OF SKILL IN THE ART

5. I previously set forth my understanding and opinions pertaining to a person of ordinary skill in the art in the relevant field (“POSA”) in my first declaration at paragraphs 60-61.

6. For example, in paragraph 61, I described that as of July 21, 2017, a POSA

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

EX1002, ¶61.

7. I understand that Dr. Mongan has provided an alternative description of the POSA as follows:

In my opinion, a person of ordinary skill in the relevant field as of July 21, 2017 would have had a bachelor’s degree or higher in computer science, bioinformatics, or a related engineering discipline, and several years of work experience relating to the development of machine learning or artificial intelligence models or algorithms. A POSA would have also worked as part of a multidisciplinary team, which would have

included individuals familiar with medical imaging, dental imaging, dentistry, and/or orthodontics.

EX2028, ¶47.

8. I disagree with Dr. Mongan's description of the person of ordinary skill in the art. For example, Dr. Mongan omits a requirement that the POSA have any experience in computer vision, image analysis, medical imaging, or related areas. The claims of the '248 patent relate to computer vision and image analysis problems; for example, the '248 patent states that it "relates to the field of dental arch image analysis." EX1001, 1:6-7. Persons of ordinary skill working on such problems would typically have the type of experience I described in my original declaration.

9. However, I understand that Dr. Mongan has also expressed that in his view, the definition he provided is not materially different from the definition I provided in my original declaration. *See* EX2002, ¶49; *see also* EX2028, ¶49.

Based on this clarification from Dr. Mongan, I interpret Dr. Mongan's characterization of the level of skill in the art as intended not to materially differ from mine. Applying this further clarification from Dr. Mongan that no material change to the level of skill I described is intended, the opinions I expressed in my original declaration would remain the same under either my description of the level of skill in the art or under Dr. Mongan's.

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

10. As I explained in my original declaration, it is my opinion that each of claims 1-2, 4, 7-8, and 13-16 was obvious in view of Kim, Invisalign, Salah, and Maninis. In this section, I address certain particular arguments raised by Drs. Mongan and Kusnoto regarding this combination.

A. The Combination of Kim, Invisalign, Salah, and Maninis is Not Inoperable

11. Dr. Mongan opines that the combination of Kim, Invisalign, Salah, and Maninis that I described in my original declaration would have been inoperable. As I discuss below, Dr. Mongan has misunderstood the teachings of the references and how I explained that a POSA would have combined them. Properly understood, the combination of the three references would have been operable.

12. To begin with, Dr. Mongan repeatedly refers to the combination as purportedly incorporating only the “low-level” parts of Maninis’s deep learning architecture, which Dr. Mongan refers to as “the Maninis COB neural network.” *E.g.*, EX2028, ¶¶125-37, 163-83, 182-99, 221-22. Dr. Mongan’s analysis is premised on a mistaken understanding that I relied solely on one part of the Maninis reference (what he calls the “COB neural network”) while not relying on Maninis’s remaining teachings of performing object classification in images.

13. In my declaration, I did not limit my reliance on the low-level portions of Maninis’s neural network architecture. On the contrary, when discussing object classification, I repeatedly cited to the discussion of high-level object classification in Maninis. For example, when discussing how “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,” I cited to both Maninis’s abstract (which discusses both low- and high-level tasks) and to pages 11-12 in section 6 of Maninis (which discusses object classification). EX1002, ¶101. In this latter section, Maninis discusses the use of neural networks that perform the task of “Semantic Boundaries” (which “requires not only detecting the boundaries, but also associating a semantic class to them”). EX1007, 11.

14. Maninis expressly teaches using the neural network and other components performing COB (as discussed earlier in Maninis) in combination with other neural network elements for object classification. EX1007, 1-2 (COB coupled to “well-known pipelines”). For example, Maninis teaches using “COB combined with the Dilated Convolution network on Semantic Boundaries,” using a combined system referred to as “COB-dil” to achieve “state-of-the-art results” in multiple metrics for semantic boundary detection. *Id.*; *see also id.*, 11-12 (discussing further implementations for performing object classification that employ COB as an element of the overall architecture, such as applications using

COB and PSPNet). I cite to Maninis’s disclosure of these high-level applications repeatedly throughout my declaration. *See, e.g.*, EX1002, ¶¶59, 75-78, 80, 97, 101-03, 123, 144, 147, 151.

15. For example, in my original declaration, I referred to Maninis’s teaching to “define boundaries around regions and to identify the objects in each region of the image” (EX1002, ¶101 (citing EX1007, Abstract, 11-12), and I explained that “[i]n 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.” EX1002, ¶102; *see also id.*, ¶124 (“[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.”) Thus, Dr. Mongan is mistaken in his assertion that I limited the proposed combination to what he calls the low-level “Maninis COB neural network” and to the exclusion of Maninis’s teaching of performing high-level tasks including object classification. EX2028, ¶183.

16. As another example, Dr. Mongan mischaracterizes my reliance on Maninis in my discussion of claim 13, asserting that “In addressing claim 13, Dr. Foroosh cites only the Maninis COB neural network, reproducing the architecture

of the Maninis COB neural network (and the subsequent contour analysis) depicted in Maninis Figure 1.” EX2028, ¶173 n.7 (citing EX1002, ¶¶146-48). This is wrong. In the cited discussion of claim 13, I cited to Maninis’s teachings of “image analysis for contour and object detection,” citing Maninis’s abstract which addresses both contour detection and object classification. EX1002, ¶147. I also referred in the same sentence back to my discussion of claim 1, which repeatedly cites to and discusses Maninis’s object classification, including the teachings relating to semantic contours. *See, e.g.*, EX1002, ¶¶97, 101-03. In discussing claim 13, I also explained that Maninis uses multiple elements, stating that it “depicts using a single-pass CNN as the *first element* of its analysis procedure,” and also citing to page 13 of Maninis, which summarizes how COB is “applied to high-level vision tasks (object proposals, object detection, and semantic contours and segmentation).” EX1002, ¶147 (citing EX1007, Abstract, 1-2, 13, Figs. 1-2); EX1007, 13.

17. Because Dr. Mongan fails to address the combination I actually discussed in my declaration, including Maninis’s teaching to use deep learning when performing object classification, his conclusion of inoperability derives from his own misunderstanding, not a problem with the art or the combination I discussed.

B. A POSA Would Have Had Reason to Use Salah and Maninis’s Deep Learning Techniques to Detect Distances Between Teeth

and Corresponding Aligner Edges with a Reasonable Expectation of Success

18. In my original declaration, I described how the remote orthodontic treatment teachings of Kim (EX1005) would be enhanced by applying Invisalign's teaching to assess whether or not an aligner was seated poorly based on a comparison with a threshold (1 mm), which Invisalign illustrates with an example image of an aligner worn on a patient's teeth, and which Invisalign characterizes as "Aligner seating poorly, >1 mm." EX1006, 1; EX1002, ¶71. I also illustrated a simple example of applying an "older and less advanced edge detection technique" (Canny edge detection), which was able to detect the edges of teeth and aligners (between which such distances would be measured) despite being less advanced than techniques taught in Salah and Maninis, such as deep learning for edge detection and object classification. EX1002, ¶102-03.

19. Dr. Mongan suggests that the image selected by Invisalign to represent a poor aligner fit was "not representative of the typical aligner fit observed by dental practitioners at the time." EX2028, ¶¶166; *see id.*, ¶90. Dr. Mongan refers to Dr. Kusnoto as the basis for this opinion. *See id.* Dr. Kusnoto asserts that different images (which he says "post-date the '248 patent") would have been more "representative of what an orthodontist would have observed in" 2017. EX2030, ¶¶61-63. For example, in a photo that Dr. Kusnoto claims to be more representative, he asserts that "separation cannot be assessed with precision."

EX2030, ¶63 (citing EX2063, 1); *see also id.*, ¶62 (asserting other images have “small separation [that] is difficult to detect”). Dr. Kusnoto does not describe any analysis or attempts to detect edges in any of his example images from which such separations can be assessed.

20. Dr. Kusnoto is mistaken in his assertion that the photos he describes as “more representative” could not be used to detect separations; in fact, detecting edges of teeth and aligners in the images is straightforward even with very old edge detection technology, and as I described in my original declaration, determining separation based on such edges is simple and straightforward. *See* EX1002, ¶¶102-07. By way of illustration, I have excerpted some of the images that Dr. Kusnoto asserts would have been representative of aligner fit (from EX2055, 2, and EX2063, 1) and performed edge detection on them using the same edge detection technique (Canny edge detection and bilateral filtering) I described in paragraph 103 of my original declaration, and the edges of the teeth and the aligner in these images can be resolved without difficulty. I have reproduced the results from this edge-detection process below, along with annotations pointing out the edges of teeth and aligners for some teeth of the depicted dental arches. Below are reproductions of two images taken from EX2055 (labeled as “excellent tracking” and “problem tracking”), as well as the image from EX2063, with each image followed by the results of the above-described edge detection process.

Fig. A: “Excellent Tracking” (EX2055, 2)

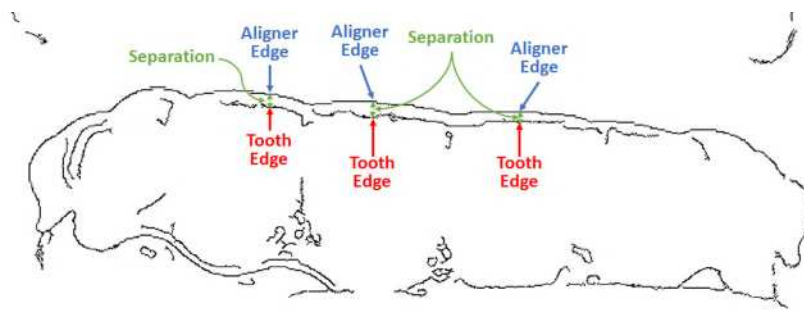


Fig. B: “Problem Tracking” (EX2055, 2)

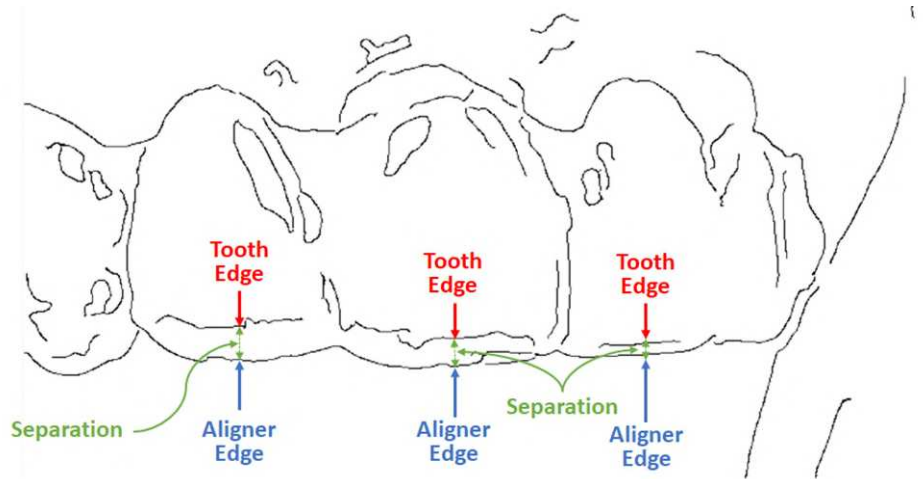
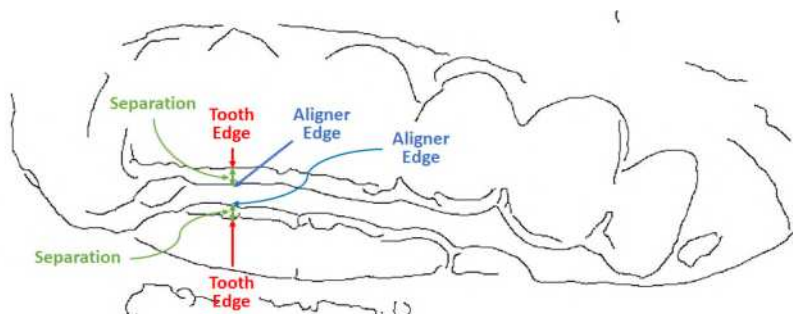


Fig. C: “Clear aligner therapy” (EX2063, 1)



21. As can be plainly seen above, the edges of teeth and aligners, and the gap between them, can be resolved even with simple edge detection techniques. This is true not only for the problem tracking image (Fig. B) and the image from the “Clear aligner therapy” reference (Fig. C), but also for the image labeled as excellent tracking (Fig. A), even though the gap between the teeth and the aligner is smaller in that image. Applying more advanced techniques, such as the deep learning taught by Salah and Maninis, would have been expected to resolve the locations of each tooth’s edge and the corresponding aligner edges with even greater precision and accuracy, allowing the detection and measurement of separations between each tooth and the corresponding edges of the aligner in a straightforward and predictable manner. Thus, contrary to Dr. Kusnoto’s assertion,

separation can be assessed in the images he asserts are representative, even using techniques less advanced than those taught by Salah and Maninis.

22. Dr. Mongan also asserts that a POSA would have been deterred from using contour detection to detect edges of teeth and aligners in images of patients wearing aligners because he claims that “Salah teaches that high contrast is required to achieve a quality analysis.” EX2028, ¶124. For this reason, Dr. Mongan asserts that a “POSA would have understood Salah to teach away from using contour detection with aligner images as the methodology would need to be altered to detect areas of low contrast instead.” *Id.*

23. Dr. Mongan is mistaken. I understand that (as Dr. Mongan acknowledges) expression of a preference for one alternative is not a teaching away; rather, to teach away a reference must “criticize, discredit, or otherwise discourage investigation into the claimed invention.” EX2028, ¶43. Salah does not teach that only maximal-contrast areas can produce useful contours. Although Salah mentions that a “blurred zone” may produce lower-quality contours (EX1008, 60:20-22), Dr. Mongan does not suggest that images of aligners on teeth would be blurred. The image from Invisalign that I analyzed in my original declaration as well as the images that Dr. Kusnoto suggests to be representative have clearly-discernable tooth and aligner edges, with no issues of blur or low contrast that would have discouraged a POSA. Indeed, Salah teaches that a range

of contrast levels may be used, noting that a threshold may be chosen to trade off between greater numbers of lower-contrast contours and smaller numbers of higher-contrast contours.

24. Furthermore, Maninis teaches how contours of different strengths may be used advantageously to discern shapes of objects (and parts of objects) at varying levels of detail. As I described in my original declaration, Maninis uses different contour strengths to detect borders between regions containing objects in the image, generating a hierarchy of regions to divide the image into its constituent objects (including subdividing objects into their constituent parts). EX1002, ¶¶59, 97, 131 (citing EX1007, 3, 5 Figs. 3-5). While even reading Salah alone would not have discouraged a POSA from applying its techniques to images of aligners worn on teeth, a POSA would have further recognized, based on the teachings of Maninis (as well as the POSA's general understanding), that any variation in contrast level in such images would not be detrimental, and in fact would be accounted for and used by Maninis's contour detection methods.

C. A POSA Would Not Have Been Deterred by Development Costs for the Proposed Combination

25. Dr. Mongan opines that the "cost, time, and complexity" of assembling and labeling training images would have deterred a POSA from practicing the proposed combination, and Dr. Kusnoto similarly opines that a POSA would not have been motivated to assemble and label images for a learning

base because he asserts that “obtaining a learning base for use in evaluating aligner fit would have been expected to be costly and complicated.” EX2028, ¶206; EX2030, ¶90; *see* EX2028, ¶¶203-06; EX2030, ¶¶85-90. I disagree.

26. Neither Dr. Mongan nor Dr. Kusnoto provides an estimate of the cost or time that they refer to. Dr. Mongan appears to estimate the costs of annotating images by assuming that POSAs would limit themselves to manual annotation of every pixel for every image in the training dataset. *See, e.g.*, EX2028, ¶205; *see also* EX2030, ¶89 (Dr. Kusnoto discussing manual labeling by persons with dental training). Certainly, manual labeling was widely practiced for labeling images by 2017 and would have been straightforward to implement with a reasonable expectation of success.

27. However, there were also numerous techniques known to POSAs in 2017 that were available to reduce the costs of generating labeled datasets, which POSAs would also have routinely employed. One such technique is data augmentation. In data augmentation, an original historical image is used to generate a set of numerous historical images by performing standard transformations on the original image. *E.g.*, EX1033, 1-2. For example, an image might be rotated to different orientations, with each rotation being used as a distinct historical image in a training dataset. The use of data augmentation was routine to persons of ordinary skill in the art and was well known in fields of

application including medical imaging; indeed, I note that Dr. Mongan was a co-author of a paper that used such techniques to generate a set of over 150,000 labeled images from 1885 labeled patient images to use in training and validation of a deep convolutional neural network. EX1033, 2. As this paper illustrates, Dr. Mongan himself was involved in a project that manually labeled over 1000 images (meeting, for example, the image-number requirements of dependent claims 7, 8, and 14), and, through the use of data augmentation, was able to further increase the size of the learning base, producing a learning base comprising more than 150,000 historical images.

28. POSAs were also familiar with other techniques used to further reduce labeling costs. One example is pseudo-labeling, in which a subset of images is initially manually labeled, and then further images are labeled with the help of a neural network trained on that subset, generating “pseudo-labels” for initially unlabeled images. *See* EX1043,¹ 2-3. The result is to train a neural network to label

¹ EX1043 is “Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks,” a paper by Dong-Hyun Lee, published on the ResearchGate website in 2015. *See* EX1043, 1 (showing submission date of August 1, 2015). This exhibit is a true and accurate copy of this paper, obtained from ResearchGate via the following download link:

images in a semi-supervised fashion, reducing the need to rely on manual labeling. *See id.*, 4-6. Further, non-deep-learning techniques were known to automatically segment images into semantically meaningful parts for easy application of labels, which can function with as few as just one labeled example. *See* EX1047,² 373-74, 381-83. Additionally, POSAs would have employed techniques to minimize the effects of any noise that arose when using automatic labeling, using techniques

https://www.researchgate.net/profile/Dong-Hyun-Lee/publication/280581078_Pseudo-Label_The_Simple_and_Efficient_Semi-Supervised_Learning_Method_for_Deep_Neural_Networks/links/55bc4ada08ae092e9660b776/Pseudo-Label-The-Simple-and-Efficient-Semi-Supervised-Learning-Method-for-Deep-Neural-Networks.pdf.

² EX1047 is “Inducing Semantic Segmentation from an Example,” a paper by Schnitman et al., published as a conference paper in 2006, in the Computer Vision—ACCV 2006 Proceedings. EX1047 is a true and correct copy of this paper, obtained from SpringerLink via a download link at https://link.springer.com/chapter/10.1007/11612704_38.

such as backward and forward correction of label noise. *See* EX1044,³ Abstract, 3-4.

29. POSAs would have appreciated that applying such routine cost-saving measures would have further lowered the costs of generating a labeled dataset of the kind I described in my original declaration, insofar as there was any concern with the costs involved.

30. Dr. Kusnoto also suggests that costs of anonymization may play a role, referring to a process of “redacting identifiable features of a patient’s face, removing patient identifying information and metadata, and assigning unique and random identifiers to the image.” EX2030, ¶88. Dr. Kusnoto does not explain what facial features would need to be redacted in the types of images at issue here. While an image showing the upper part of a person’s face, including their eyes, might be identifiable, that does not mean that images of a mouth would be identifiable. For example, the images shown in Salah Figs. 5a-6a show only a

³ EX1044 is “Making Deep Neural Networks Robust to Label Noise: a Loss Correction Approach,” a paper by Patrini et al. published on arXiv.org in March 2017. *See* EX1044 (showing submission date of March 22, 2017). This exhibit is a true and accurate copy of this paper, obtained from arXiv.org via the following download link: <https://arxiv.org/pdf/1609.03683v2>.

patient's dental arches and the lower parts of the face. EX1008, Figs. 5a-6a; *see also* EX1005, Figs. 8-13 (similar images in Kim). Regarding removing patient identifying information and metadata from image files, as well as assigning random identifiers, these processes can easily be performed automatically at practically no cost, using programs or scripts that would be elementary for a POSA to write.

31. Regardless, a POSA would not have been deterred even by substantial development costs. The production of a learning base is essentially a one-time, upfront expense as part of the development process of a deep learning device. A POSA would not have been discouraged by the prospect of development costs of tens or hundreds of thousands of dollars. Indeed, even costs over \$1 million would not have been seen by a POSA as a significant deterrent to development of such a learning base. Such costs are not particularly high even compared to typical budgets for academic research, to say nothing of the vastly larger budgets available for development of commercially-useful products. For example, my CV details research projects with funding totaling over \$14 million, including multiple multi-million-dollar projects, and I note that Dr. Mongan's CV reports similar budgets for certain of his own funded development projects. *See* EX1003, 9-13; EX2029, 4, 28-32. Thus, a POSA would be motivated to combine Kim, Invisalign, Salah, and

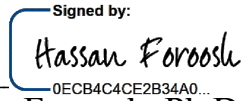
Maninis as described in my original declaration even in the face of substantial associated costs.

V. CONCLUDING STATEMENTS

32. 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.

33. 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 4, 2024

By:  Signed by:
Hassan Foroosh, Ph.D.
0ECB4C4CE2B34A0...

VI. APPENDIX – CITED PETITIONER EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 11,049,248 to Salah et al.
1002	Declaration of Hassan Foroosh, Ph.D.
1003	Curriculum Vitae of Hassan Foroosh, Ph.D.
1005	U.S. Patent Pub. No. 2017/0281313 to Taewon Kim (“Kim”)
1006	Align Technology, Invisalign Clinical Monitoring Guide (2003) (archived September 26, 2003, at http://web.archive.org/web/20030926024959/http://www.invisaligncec.com:80/consistent/pdfs/ICMG_Final.pdf) (“Invisalign”)
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.
1033	Rajkomar et al., “High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks,” <i>J. Digital Imaging</i> , 30, 95-101 (Feb. 2017)
1043	Dong-Hyun Lee, “Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks,” <i>ResearchGate</i> (https://www.researchgate.net/profile/Dong-Hyun-Lee/publication/280581078_Pseudo-Label_The_Simple_and_Efficient_Semi-Supervised_Learning_Method_for_Deep_Neural_Networks/links/55bc4ada08ae092e9660b776/Pseudo-Label-The-Simple-and-Efficient-Semi-Supervised-Learning-Method-for-Deep-Neural-Networks.pdf) (Aug. 1, 2015)
1044	Patrini et al., “Making Deep Neural Networks Robust to Label Noise: a Loss Correction Approach,” <i>arXiv.org</i> (https://arxiv.org/pdf/1609.03683v2) (Mar. 22, 2017)

1047	Schnitman et al., “Inducing Semantic Segmentation from an Example,” <i>Computer Vision—ACCV 2006</i> (2006)
------	---

VII. APPENDIX – CITED PATENT OWNER EXHIBITS

Exhibit No.	Description
2002	Declaration of Dr. John Mongan in Support of Patent Owner's Preliminary Response
2028	Declaration of Dr. John Mongan in Support of Patent Owner's Response
2029	Updated Curriculum Vitae of Dr. John Mongan
2030	Declaration of Dr. Budi Kusnoto in Support of Patent Owner's Response
2055	Making Sure Your Invisalign Aligners Fit Correctly, Morita Orthodontics
2063	N. Smith, Clear aligner therapy: An esthetic approach to total patient care, Dentistry IQ