

Petitioner's Demonstratives

Align Technology, Inc. v. Dental Monitoring SAS

IPR2023-01369

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

Instituted Ground

Claims	35 U.S.C. §	Prior Art
1-15	103	Salah, Carrier, and Maninis

The '409 Patent

(57)

ABSTRACT

A method for acquiring an image of a dental arch of a patient. Activation of an image acquisition apparatus so as to acquire an analysis image of the arch. Analysis of the analysis image by a deep learning device trained by a learning base. Determination, for the analysis image, as a function of the results of the analysis, of a value for an image attribute. Comparison of the image attribute value with a setpoint. Sending of an information message as a function of the comparison, the information message being related to the quality of the image acquired and/or to the position of the acquisition apparatus in relation to said arch and/or to the setting of the acquisition apparatus and/or to the opening of the mouth and/or to the wearing of a dental appliance.

The invention proposes a method for analyzing an image, called “analysis image”, of a dental arch of a patient, a method in which the analysis image is submitted to a deep learning device, preferably a neural network, in order to determine at least one value of a tooth attribute relating to a tooth represented on the analysis image, and/or at least one value of an image attribute relating to the analysis image.

A “neural network” or “artificial neural network” is a set of algorithms well known to a person skilled in the art.

The neural network may in particular be chosen from: the networks specializing in the classification of images, called “CNN” (“convolutional neural network”), for example

AlexNet (2012)

ZF Net (2013)

VGG Net (2014)

GoogleNet (2015)

Microsoft ResNet (2015)

Caffe: BAIR Reference CaffeNet, BAIR AlexNet

Torch: VGG_CNN_S, VGG_CNN_M,
VGG_CNN_M_2048, VGG_CNN_M_10 24,
VGG_CNN_M_128, VGG_CNN_F, VGG ILS-
VRC-2014 16-layer, VGG ILSVRC-2014 19-layer,
Network-in-Network (Imagenet & CIFAR-10)

Google: Inception (V3, V4).

The networks specializing in the location and detection of objects in an image, the object detection network, for example:

R-CNN (2013)

SSD (Single Shot MultiBox Detector: Object Detection network), Faster R-CNN (Faster Region-based Convolutional Network method: Object Detection network)

Faster R-CNN (2015)

SSD (2015).

The above list is not limiting

Claim 1

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

- a') activation of an image acquisition apparatus so as to acquire an image, called “analysis image”, of said arch;
 - b') analysis of the analysis image by means of a deep learning device trained by means of a learning base;
 - c') determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute;
 - d') comparison of said image attribute value with a setpoint;
 - e') sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof,
- to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image.

Claim Construction: Petition

Term	Petition	PO response
image attribute	encompasses attribute relating to image as a whole	No response
tooth attribute	any attribute related to a tooth	No response
analysis of the analysis image by means of a deep learning device trained by means of a learning base	function: analyzing an image structure: a device comprising a deep learning neural network trained with a learning base	No response

Pet. 15-18

Claim Construction: Patent Owner

Term	POR	Effect
deep learning device	a computer program or algorithm	No effect on petition grounds
real time (claim 14)	occurring while the image acquisition process is still underway	No effect on petition grounds
tooth shape parameter (claim 4)	not a tooth contour	Not argued to be relevant in POR

POR 13-17; Reply 2-4; EX1037, 62:10-63:11

Salah

(54) Title: METHOD FOR MONITORING DENTITION

(72) **Inventors:** Salah, Philippe; 18, rue Edouard Vaillant, F-93170 Bagnolet (FR). Ayache, William; 18, rue Ancelle, F-92200 Neuilly-sur-Seine (FR). Debraux, Laurent; 57, rue des Cascades, F-75020 Paris (FR) Ghy-Selinck, Guillaume; 41, rue d'Arleux, F-59169 Cantin (FR). Pellissard, Thomas; 5, rue de la famille Solomon, F-95190 Goussainville (FR).

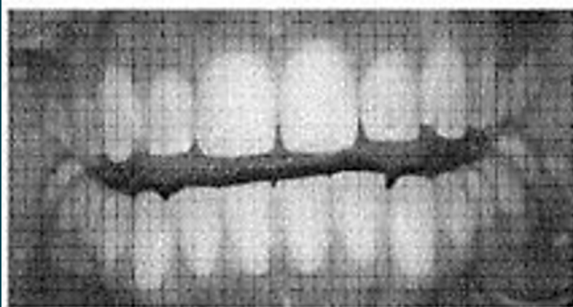


Fig. 6a

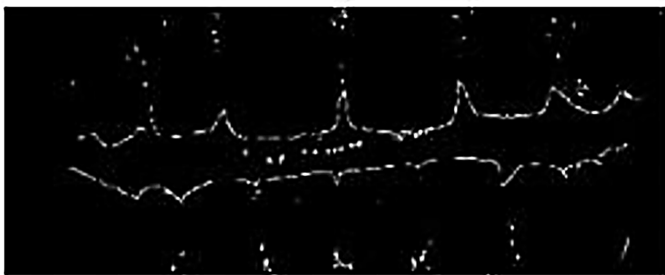


Fig. 6b

Preferably, however, the image-acquisition apparatus includes verification means to facilitate its approximate positioning relative to the patient prior to acquisition of the updated image.

The user can be guided through the acquisition process by written and/or voice prompts. By

For example, the personal device may announce “take a photo from the front”, emit a signal to inform the user that the photo is acceptable or that, on the contrary, he/she must take another photo, announce “take a photo from the right”, preferably displaying an arrow to orient the user, etc. The completion of the acquisition process may also be announced by the device. The device can also help with positioning, for example by visual prompts (e.g. by displaying arrows), and/or audio messages (such as a series of beeps, the frequency of which increases as the device's positioning improves), and/or written and/or vocal prompts (“higher”, “lower”, etc.).

In particular, verification means can include references that appear on the viewfinder or screen. The references may, for example, comprise a line intended to be aligned with the general direction of the juncture between the upper and lower teeth when the teeth are clenched by the patient, and/or a vertical line intended to be aligned with the joint between the two upper incisors. The references can also refer to other parts of the patient. For example, they may take the form of marks corresponding to the position of the eyes, or take the form of a shape in which the patient's mouth or face is to be positioned.

provided by image processing software. Such processes can be used, for example, to detect high-contrast regions and determine contours.

- by “Deepleraning” [sic].

EX1008, [54], [72], 27:19-28:10, 34:8-36:14; Figs. 6a-b;
EX1012

Carrier

METHODS AND APPARATUSES FOR DENTAL IMAGES

first color to a second color (e.g., green, blue, etc.) and/or automatically taking the image. Note that the method or apparatus may automatically determine alignment between the overlay and the patient's teeth in a variety of ways. For example, the method or apparatus may determine alignment when a threshold amount/percentage of the edges of the patient's teeth in the image are within a predetermined minimum distance of the overlay. For example, alignment may be indicated when greater than 50%, greater than 60%, greater than 70%, greater than 80%, greater than 90%, etc. of the edges of the teeth are within about 1 mm, about 0.5 mm, about 0.1 mm, about 0.05mm, etc.). The edges of the teeth may be determined by any appropriate technique (e.g., edge detection image processing, etc.).

[0081] The method can comprise evaluating the image quality and warning doctors if the image might be rejected. For example, if the image is too dark, or blurry, or out of focus, a warn message is shown on the screen to warn the doctors to retake the photo. The quality of the image can be evaluated and reviewed by a variety of methods. For example, the method can comprise evaluating the focus status of the image. For another example, the method can comprise analyzing the image using a library. For yet another example, the method can comprise analyzing the brightness of the image. In some embodiments, the method can further comprise using image recognition algorithms to ensure that all required teeth are visible.

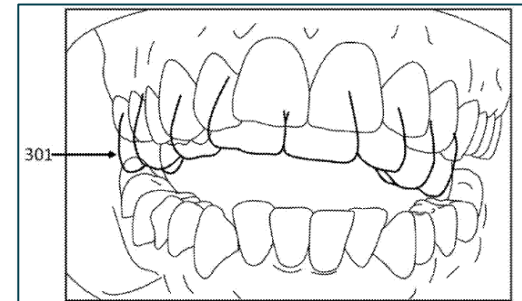


FIG. 3A

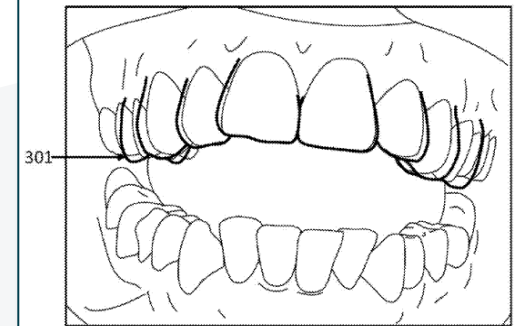


FIG. 3B

EX1005, Title, ¶¶108, 128, Figs. 3A-B; EX1006, Title, ¶¶63, 81; Figs. 3A-3B

Maninis

Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks

Abstract—We present Convolutional Oriented Boundaries (COB), which produces multiscale oriented contours and region hierarchies starting from generic image classification Convolutional Neural Networks (CNNs). COB is computationally efficient, because it requires a single CNN forward pass for multi-scale contour detection and it uses a novel sparse boundary representation for hierarchical segmentation; it gives a significant leap in performance over the state-of-the-art, and it generalizes very well to unseen categories and datasets. Particularly, we show that learning to estimate not only contour strength but also orientation provides more accurate results. We perform extensive experiments for low-level applications on BSDS, PASCAL Context, PASCAL Segmentation, and NYUD to evaluate boundary detection performance, showing that COB provides state-of-the-art contours and region hierarchies in all datasets. We also evaluate COB on high-level tasks when coupled with multiple pipelines for object proposals, semantic contours, semantic segmentation, and object detection on MS-COCO, SBD, and PASCAL; showing that COB also improves the results for all tasks.

At the core of all these deep learning approaches lies a *base CNN*, starting from the seminal AlexNet [8] (8 layers), through the more complex VGGNet [10] (16 layers) and inception architecture of GoogLeNet [9] (22 layers), to the very recent and very deep ResNets [11] (up to 1001 layers). Image classification results, which originally motivated these architectures, have been continuously improved by exploring deeper and more complex networks. In this work, we present results both using VGGNet and ResNet, showing that COB is modular and can incorporate and benefit from future improvements in the base CNN.

of the art. We have also studied the effects of COB when coupled with well-known pipelines, showing that injecting COB detections to them lead to improvements on Semantic Segmentation and Object Detection. Finally, we report a new state of the art on Semantic Boundary detection.

EX1007, Abstract, 2, 13; see also *id.*, 11-12, Tables 4-7

Maninis

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).

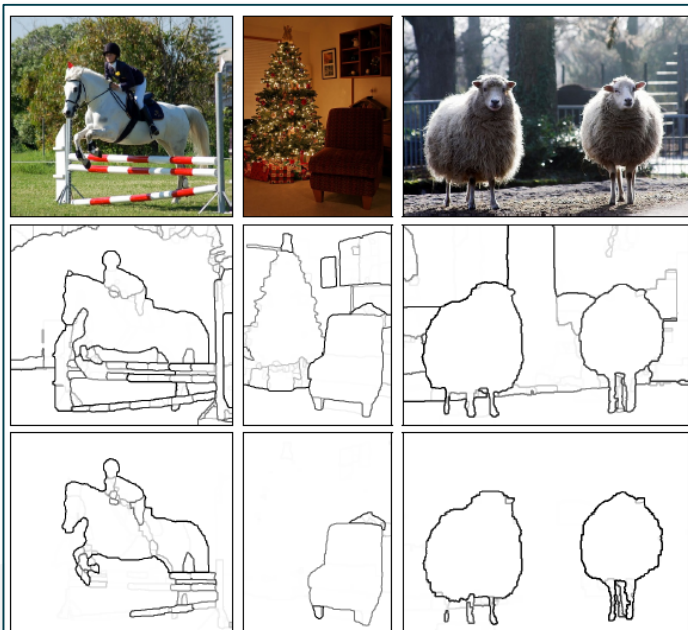


Fig. 10. **Qualitative results for Object Boundaries.** Row 1: original images, Row 2: Generic Image Segmentation results, Row 3: Object Boundary results.

Our technique achieves state-of-the-art performance by a significant margin for contour detection, the estimation of their orientation, and generic (RGB and RGB-D) image segmentation. We show that our architecture is modular by using two different CNN base architectures, which suggests that it will be able to transfer further improvements in CNN base architectures to perceptual grouping. We also show that our method does not require globalization, which was a speed bottleneck in previous approaches. The generalization of COB was further demonstrated when applied to high-level vision tasks (object proposals, object detection, and semantic contours and segmentation) in combination with recent pipelines, where the results are improved in all cases.



Fig. 16. **Qualitative results for Semantic Segmentation.** Row 1: original images, Row 2: Dilated Convolution Network, Row 3: Dilated Network with COB superpixels.

EX1007, 9, 11-13, Figs. 10, 16; *see also id.*, Tables 4-7

Proposed Combination

- Acquisition of dental images, as taught by Salah;
- Guided using the edge detection and object recognition of Carrier to align overlay and ensure all required teeth are visible;
- Implemented using deep learning device as taught by Maninis

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

EX1002, ¶86.

images with users' mobile phones. Both references teach assessing alignment and using the assessment to guide users, with Carrier providing specific details regarding how to effectively assess alignment based on the appearance of the patient's teeth, thereby ensuring high-quality dentition images. A POSA would thus have considered Carrier's teachings directly applicable to the types of imaging methods taught by Salah and would use Carrier's complementary teachings to supplement Salah's disclosure. For example, a POSA would be motivated to use Carrier's measures of image alignment and the presence of all required teeth to provide alignment guidance to users as well as to assess the acceptability of alignment based on images being acquired during the alignment process.

Carrier is Prior Art

- Petition relies on material undisputedly taught in Carrier Provisional.
- Under *Penumbra, Inc. v. RapidPulse, Inc.*, IPR2021-01466, Paper 34 at 32 (Mar. 10, 2023) (precedential, §II.E.3), Carrier is prior art.
- PO argues that Carrier Provisional does not support Carrier Publication’s claim elements of
 - “a mobile application on a mobile device associated with a patient”; and
 - “guiding a patient.”
- PO admits Carrier Provisional teaches these elements for a cell phone’s *user*, but argues that the Carrier Provisional only discloses *dental professionals* taking photos

Carrier is Prior Art

Dr. Kusnoto:

75. Dr. Foroosh cites instances in which the Carrier Provisional uses the term “user” in a context where “user” implies “patient” but not a dental professional. EX1002, ¶¶49, 55 (citing EX1006, [0055] (indicating overlay can be derived from “user’s teeth,” referring to a patient’s and not a dentist’s teeth)). However, in other

Q. But it doesn't say that in those paragraphs. I'm not asking -- I'm trying to set aside the issue of who that user is, and I'm focused on what the user does.

A. The user takes the pictures.

Carrier Provisional:

[0055] An overlay may include an outline (e.g., a perspective view outline) of a set of teeth that may be used as a guide to assist in placement of the camera to capture the patient image. The overlay may be based on a generic image of teeth, or it may be customized to the user’s teeth, or to a patient-specific category (by patient age and/or gender, and/or diagnosis, etc.). The overlay may be shown as partially transparent, or it may be solid, and/or shown in outline.

[0062] In some other embodiments, the overlay can be a patient-specific overlay derived from the patient’s teeth. In some variations the apparatus (e.g., software/application software) may image the users teeth or may receive data derived from a scan, including digital scan, of the users teeth, and may generate an overlay, such as the outline overlay shown in FIGS. 2A-2C. For example, the patient-specific overlay 301 can be obtained from a specific model of teeth for a specific patient and/or from 3D scan or 3D impressions of the teeth of the patient. This may provide optimal matching for positioning the handheld electronics (e.g., smart phone) when taking the image(s).

Patients Can Take Carrier Provisional's Images

Dr. Kusnoto:

high-quality dental photographs. In fact, **this standard set includes photographs that a patient cannot (or cannot easily) take of themselves**, e.g., the occlusal views (shown in Figure 5A) and the external head or facial views (shown in Figures 5C and D). *Id.*, [0076]-[0077].

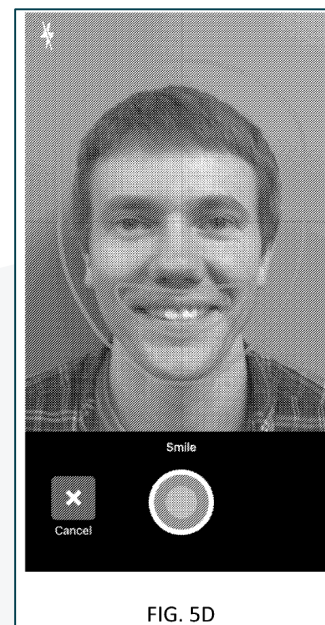
Dr. Valley:

EX2030, ¶55. In fact, the Carrier Provisional discloses many angles that a patient could easily take of themselves. Dr. Kusnoto says that the image in Fig. 5D would be difficult or impossible for a patient to take (EX2030, ¶55), but Fig. 5D of the Carrier Provisional shows an ordinary picture of a smiling face that would be captured in a typical, easily taken “selfie.” See EX1006, Fig. 5D. Similarly, the

allows for images from custom angles. EX1006, ¶¶75-78. A dental professional reading the Carrier Provisional would have appreciated that in many cases it would be useful to **obtain fewer than all of the images depicted in the figures** and would not have interpreted the disclosure of the Carrier Provisional to be restricted to capturing each and every one of the images shown in the figures.

Carrier Provisional:

[0075] For example, a method to obtain an image of the teeth may include displaying, on the screen, an overlay comprising an outline of teeth in one of an anterior view, a buccal view, an upper jaw view, or a lower jaw view, wherein the overlay is displayed atop the view of the patient's teeth. The method and apparatus may **walk the user through taking a complete or partial set of these images**. For example, FIG. 5A shows an example of a user interface to enable



Reply 4-5; EX1035, ¶¶32-33; EX2030, ¶55; EX1006, ¶75, Fig. 5D.

Maninis is Prior Art

- No dispute on date of submission for publication (months before earliest claimed priority date of July 21, 2017):

arXiv:1701.04658v2 [cs.CV] 28 Apr 2017

- No dispute materially identical version (EX1021) published prior to April 28, 2017. *See Valve Corp. v. Ironburg Inventions Ltd.*, 8 F.4th 1364, 1370-71 (Fed. Cir. 2021).
- Archived policy demonstrates publication no later than May 2, 2017:

Availability of submissions and announcement order		
Note: policy change 2017-01 regarding announcement order and the effect of edits.		
Submissions to arXiv are normally made public on Sunday through Thursday, with no announcements Friday or Saturday. Check the current time at the arXiv main site, with the amount of time remaining before the "freeze".		
Submissions Received Between (all times Eastern US)	Will Be Announced (all times Eastern US)	Mailed to Subscribers
Monday 14:00 - Tuesday 14:00	Tuesday 20:00	Tuesday Night / Wednesday Morning
Tuesday 14:00 - Wednesday 14:00	Wednesday 20:00	Wednesday Night / Thursday Morning
Wednesday 14:00 - Thursday 14:00	Thursday 20:00	Thursday Night / Friday Morning
Thursday 14:00 - Friday 14:00	Sunday 20:00	Sunday Night / Monday Morning
Friday 14:00 - Monday 14:00	Monday 20:00	Monday Night / Tuesday Morning

- *Advanced Micro-Devices, Inc.*, 848 F.2d 1560, 1569 (Fed. Cir. 1988): “Evidence of routine business practice can be sufficient to prove that a reference was made accessible before a critical date.”

Pet., 10-11; EX1007, 1; EX1002, ¶¶62-67; EX1021; EX1022, 3; Reply, 6; EX1023; EX1024; EX1025; Paper 35

Rationale to Combine

- Salah and Carrier provide complementary teachings to acquire dental images
 - Carrier provides specific details regarding how to effectively assess alignment, directly applicable to supplement Salah's imaging methods
 - Carrier's techniques expected to ensure high-quality images
 - Predictable combination, with Carrier providing detailed methods to guide users and assess image acceptability to acquire dental images as taught by Salah
- Further combination with Maninis provides state-of-the-art edge and object detection to accomplish overlay alignment and to ensure required teeth are visible
 - Maninis's techniques broadly applicable and expected to improve performance
 - Maninis teaches its techniques represent current trend in computer vision
 - Carrier openly invites use of known edge detection and object recognition techniques
 - Salah teaches use of deep learning for detecting "discriminant information" including tooth edges
- Combination would "adapt Salah and Carrier's image acquisition device to employ a deep learning architecture as taught by Maninis, trained on images relevant to Salah's and Carrier's dental imaging applications," with reasonable expectation of success

Salah Not Superior to Carrier

- PO argues, citing Salah's "registration marks," that "Carrier's technique does not provide the 'same function' as Salah's—Carrier's is significantly worse, and its adoption would have been detrimental to the operation of Salah"
- This ignores Salah's teachings:

Preferably, however, the image-acquisition apparatus includes verification means to facilitate its approximate positioning relative to the patient prior to acquisition of the updated image.

The user can be guided through the acquisition process by written and/or voice prompts. By

For example, the personal device may announce "take a photo from the front", emit a signal to inform the user that the photo is acceptable or that, on the contrary, he/she must take another photo, announce "take a photo from the right", preferably displaying an arrow to orient the user, etc. The completion of the acquisition process may also be announced by the device. The device can also help with positioning, for example by visual prompts (e.g. by displaying arrows), and/or audio messages (such as a series of beeps, the frequency of which increases as the device's positioning improves), and/or written and/or vocal prompts ("higher", "lower", etc.).

In particular, verification means can include references that appear on the viewfinder or screen.

The references may, for example, comprise a line intended to be aligned with the general direction of the juncture between the upper and lower teeth when the teeth are clenched by the patient, and/or a vertical line intended to be aligned with the joint between the two upper incisors. The references can also refer to other parts of the patient. For example, they may take the form of marks corresponding to the position of the eyes, or take the form of a shape in which the patient's mouth or face is to be positioned.

- PO concedes "Carrier further teaches that such an overlay can be used to capture high quality dental photographs...."

Pet., 10, 21; EX1002, ¶¶49, 81; POR, 46, 58; Reply, 11

Salah Not Superior to Carrier

- Registration marks would move substantially relative to the teeth, making them less reliable than alignment to teeth themselves.

Dr. Kusnoto:

kind of things, it depends, but generally the tooth movement is within half to a millimeter a month.

Dr. Valley:

42. As I discussed above, Carrier teaches using patient- and treatment-specific overlays when aligning to patients' teeth. See EX1006, ¶¶55, 62, 64-65; EX1005, ¶¶100, 107, 109-10. To the extent there was a concern regarding variation between patients in tooth size or position, Carrier's use of customized patient-specific overlays and treatment-specific overlays would eliminate such a concern. This is because, a patient-specific overlay or a treatment-specific overlay would reflect the patient's specific dental anatomy, including differences of a specific

by a patient, the positions of registration marks like the ones shown in Salah or in Dr. Kusnoto's declaration will move relative to the patient's teeth. Further, the position of the registration marks will depend on the way the retractor is worn—if a patient inserts the retractor at an angle, the registration marks at the edges of the retractor will move significantly relative to the patient's teeth. Unlike expected movement of teeth, these types of registration marks would be expected to move unpredictably over distances much more than a millimeter. Where the purpose is to produce images of a patient's teeth, the teeth themselves would provide a better reference for alignment than registration marks, which can move substantially relative to the teeth to be imaged.

POR, 47; Reply, 12-13; EX1035, ¶¶43-44; EX1038, 78:10-21

Salah Not Superior to Carrier

- Registration marks provide fewer pixels, reducing alignment reliability.

Dr. Foroosh:

38. Dr. Mongan's proposed use of registration markers would also have drawbacks relative to Carrier's alignment techniques. For example, the registration marks visible on Dr. Kusnoto's retractor image are very small, so will appear in an image as a relatively small number of pixels, with relatively small areas and perimeters. See EX2030, ¶37. Such a marker provides limited reliability in determining position and orientation of the imaging device in comparison to larger reference objects such as the contours of teeth that provide many more pixels for image alignment and registration. *Compare id.* (registration marks small compared

attached to retractors placed in a patient's mouth. Salah's registration marks are thus only indirectly related to the position of the patient's teeth. When seeking to capture an image of an object, aligning to the object itself is a more straightforward method than aligning to other objects. As a matter of common sense, if you want to capture an image of certain teeth, a natural way to ensure success is to detect the teeth themselves.

Reply, 12-13; EX1035, ¶42; EX1038, 78:10-21

Combination with Maninis Works as Proposed

- PO argues: “Maninis COB neural network is incapable of determining what [depicted] objects might be”
- Dr. Mongan admits he uses the term “COB neural network” to exclude Maninis’s teachings of using COB in a pipeline for object classification (e.g., in Maninis section 6, pp. 11-12)
- The petition proposed using Maninis’s deep learning architecture to detect contours and also to classify objects

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

applications. Maninis’s deep learning architecture would provide specific functionalities required by Salah and Carrier, including the ability to identify objects (such as the required teeth Carrier proposes using image recognition software to identify) and to detect contours (such as the tooth edges Carrier uses to assess alignments), to aid in guiding image acquisition, as taught by Salah.

Maninis’s deep learning algorithm uses the contours it generates to define boundaries around regions and to identify the objects in each region of the image. See EX1007, Abstract, 11-12. Maninis describes such image recognition tasks as involving “image classification, object detection, and semantic segmentation” where detected contours are assigned to categories matching the “ground-truth” (i.e., identification what actual objects are where in an image). EX1007, 1-2.

Pet., 23-24, 33-34; EX1002, ¶¶83, 85-86, 117; Reply, 7-8; EX1048, ¶¶23-28; EX1037, 104:11-19, 107:2-17, 103:14-104:10, 118:13-119:11; POR, 41

POSAs Motivated to Use Deep Learning: PO's Speed Concerns Meritless

- PO argues deep learning “analysis could take more than 6 seconds on a mobile device,” or “could easily take twice as long,” which “would have precluded real time feedback to guide the user on their mobile phone, effectively rendering those alignment techniques unusable”
- However, the claims separately recite a “deep learning device” and an “image acquisition device”—nowhere do the claims require the deep learning device to be run on a mobile device
- As the institution decision recognized, the petition did not limit the combination to a deep learning device (e.g. neural network architecture) running on a cell phone

Pet., 30:

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

Pet., 31:

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

ID, 32:

18:56–59, 23:7–10. The record before us is devoid of any evidence to substantiate Patent Owner's inference that the “deep learning device” recited in the claim is somehow integrated in the mobile device. Accordingly, on this record, we find insufficient evidence to corroborate Patent Owner's argument that the claimed “deep learning device” is somehow different from the one disclosed in Maninis.

See also EX1048, ¶¶41-42; POR, 53-54

POSAs Motivated to Use Deep Learning: PO's Speed Concerns Meritless

- Petition presented mappings with two distinct types of image attributes, each of which would individually satisfy all claim elements (edge-overlay distances and presence of all required teeth)
 - PO's slow-processing arguments fail to address required-teeth mapping, which provides feedback to re-take photos
- POSAs had ample tools to implement neural networks on mobile devices that provide fast processing
- Image classifiers taught by Maninis used only minimal time (e.g., 60 ms/image for Faster R-CNN)
- POSAs accounted for increasing computing power when developing software

Reply, 16-18; EX1048, ¶¶43-44, 46-47; *see also* Pet., 34; EX1002, ¶118 (adjusting Maninis's neural network straightforward based on POSA's skill)

No Efficiency Improvement in '409 Patent

'409 patent relies on known, generic deep learning:

A "neural network" or "artificial neural network" is a set of algorithms well known to a person skilled in the art.

The neural network may in particular be chosen from: the networks specializing in the classification of images, called "CNN" ("convolutional neural network"), for example

AlexNet (2012)

ZF Net (2013)

VGG Net (2014)

GoogleNet (2015)

Microsoft ResNet (2015)

Caffe: BAIR Reference CaffeNet, BAIR AlexNet

Torch: VGG_CNN_S, VGG_CNN_M,

VGG_CNN_M_2048, VGG_CNN_M_10 24,

VGG_CNN_M_128, VGG_CNN_F, VGG ILS-

VRC-2014 16-layer, VGG ILSVRC-2014 19-layer,

Network-in-Network (Imagenet & CIFAR-10)

Google: Inception (V3, V4).

The networks specializing in the location and detection of objects in an image, the object detection network, for example:

R-CNN (2013)

SSD (Single Shot MultiBox Detector: Object Detection network), Faster R-CNN (Faster Region-based Convolutional Network method: Object Detection network)

Faster R-CNN (2015)

SSD (2015).

The above list is not limiting

Maninis improves over state-of-the-art:

(Column (3)). Our final technique COB, in which we remove the globalization step, computes the three scales in one pass and add contour orientations, takes 0.79 seconds in mean. 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).

Our technique achieves state-of-the-art performance by a significant margin for contour detection, the estimation of their orientation, and generic (RGB and RGB-D) image segmentation. We show that our architecture is modular by using two different CNN base architectures, which suggests that it will be able to transfer further improvements in CNN base architectures to perceptual grouping. We also show that our method does not require globalization, which was a speed bottleneck in previous approaches. The generalization of COB was further demonstrated when applied to high-level vision tasks (object proposals, object detection, and semantic contours and segmentation) in combination with recent pipelines, where the results are improved in all cases.

EX1007, Abstract, 9, 13; EX1001, 16:15-43; Reply/Pet.

No Deterrence from Training Costs

- Dr. Mongan admits estimates based on unneeded steps:

Q. And were you labeling just the edges of the teeth and the tooth numbers or was there additional labeling?

A. There are some additional parameters like inter-molar width, inter-canine width, the depth of the arch, the contour of the tongue, a few other parameters.

Q. Did that project involve labeling any issue with the teeth, like malocclusion?

A. Could be, yeah. Could be. The one with the American Board of Orthodontics, I think it is, yeah.

Q. Right.

And if it's an image of just of the teeth, not the patient's face, then you don't need to redact the image; right?

MR. BLYTHE: Objection to form.

THE WITNESS: The image itself, no, but the identifier that goes with the image, yes.

Q. And when you talk about the time that it would take to label and your estimate of that and the 20 minutes that you think it would take to label, that's based on your own experience that you just described earlier in that project from 2016; right?

A. Yeah, when the -- as I said, there are multiple times we are doing similar things. So the number always came up would be about the same per image.

- Dr. Foroosh:
Costs not an issue regardless:

deep learning device. Costs in the tens to hundreds of thousands of dollars are quite small compared to typical development costs even of non-commercial products, and a POSA would not have been discouraged by the prospect of such development costs.

56. Dr. Kusnoto disagrees with Dr. Mongan's cost estimate, suggesting that costs may rise as high as \$1.5 million, including as much as 2 years of dental student labor. EX2030, ¶49. This cost, too, 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

Reply, 19-20; EX1038, 81:20-25, 88:9-17, 91:9-18; EX1049, ¶¶50-56

Carrier Teaches Detecting Required Teeth

Carrier:

[0081] The method can comprise evaluating the image quality and warning doctors if the image might be rejected. For example, if the image is too dark, or blurry, or out of focus, a warn message is shown on the screen to warn the doctors to retake the photo. The quality of the image can be evaluated and reviewed by a variety of methods. For example, the method can comprise evaluating the focus status of the image. For another example, the method can comprise analyzing the image using a library. For yet another example, the method can comprise analyzing the brightness of the image. In some embodiments, the method can further comprise using image recognition algorithms to ensure that all required teeth are visible.

Petition:

¶135. The messages also relate to image quality and opening of the mouth. Salah expressly teaches messages regarding whether the image is “acceptable,” which is an assessment of quality. EX1008, 27:21-28:2; EX1002, ¶136. Where acceptability was based on the presence or absence of required teeth, as taught by Carrier, this would relate to quality (lack of visible teeth indicating poor quality), positioning (as mis-positioning can result in required teeth not being in frame), and opening of the mouth (as a closed mouth would obscure teeth). EX1002, ¶137; *see, e.g.*, EX1006, Figs. 9-10 (showing teeth no longer visible when mouth is closed); EX1005, Figs. 9-10.

Dr. Valley:

39. From the point of view of a dental professional, it is easy to appreciate that even if a camera is perfectly aligned to a particular predetermined viewpoint, it may not capture all the teeth that are intended to be visible from that viewpoint. Notwithstanding perfect alignment, some teeth may be obscured by objects, such as the patient’s tongue or lips, or a retractor that has been positioned at an incorrect angle, or a hand or finger of the operator mistakenly inserted in front of the camera

109-10. If a patient had missing or unerupted teeth, this fact would be indicated in dental records, including the scans discussed by Carrier (*see* EX1006, ¶62; EX1005, ¶107), and thus the patient-specific overlay would indicate the absence of those teeth, meaning that the “required teeth” for such a patient would *not* include teeth the patient didn’t have (or that were not visible as they were unerupted).

Pet., 41; EX1006, ¶81; EX1005, ¶128; EX1035, ¶¶39-40; *see also* Reply, 20-22; EX1048, ¶¶47-63; EX1002, ¶137.

Claim 7

...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”...

- Every image of teeth has attribute values for image attributes, such as the set of teeth depicted in the image
- Claim recites no requirements regarding presence or manner of labels
- Pixelwise labeling of tooth numbers for each depicted tooth is a labeling of an image attribute value
- Image attribute value can be a set of depicted teeth:

In the step c'), an attribute value of the analysis image is determined as a function of the tooth attribute values. The attribute value of the analysis image may relate to its general orientation and may for example take one of the following three values: “right photo”, “left photo” and “front photo”. The attribute value of the analysis image may also be the list of the tooth numbers represented, for example, “16, 17 and 18”. The attribute value of the analysis image may also be, for example, the “presence” or “absence” of a dental, preferably orthodontic, appliance, or the state of opening of the mouth (“mouth open”, “mouth closed”).

Pet., 15-16, 50-52; Reply, 23-24; EX1002, ¶¶73-74, 183-211; EX1001, 26:16-26

Claims 8-9

- PO: “Petitioner identifies no motivation to” practice the claimed step of “transferr[ing] information from the 3D model to the updated images,” however:
 - PO argues lack of express disclosure but ignores the petition’s obviousness case, showing it was obvious to label using tooth numbers
 - PO has no response to evidence showing labeling of images using 3D models
 - PO faults petition for not providing a reason for doing what Salah expressly teaches—acquiring images and matching them to 3D models

Pet., 52-65; Reply, 24-26; EX1002, ¶¶183-211;
EX1008, 23:4-8, Fig. 4d; EX1013, 3689-90; EX1014, 3198, 3200-01

Claim 14

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

- Salah and Carrier expressly teach checking and feedback in real time
- Petition proposed combination providing feedback during image acquisition
- Dr. Mongan conceded PO's construction of real time is satisfied so long as feedback is received during image acquisition:

A So "immediately" is somewhat context-dependent, but it means soon enough to be part of the acquisition process and for the analysis and feedback that is provided to be useful in the context of that acquisition process.

Q So you're not putting a particular time duration limitation on that; is that right?

A I'm not generally establishing any specific time duration to --

The approximate time duration would be somewhat context-dependent.

Pet., 68-70; EX1006, ¶63; EX1005, ¶108; EX1008, 27:21-28:2; EX1002, ¶¶227-28; EX1048, ¶17; EX1037, 62:10-63:11; Reply, 3, 26

Motion to Exclude Should be Denied

- EX1031:
 - Timely
 - No evidentiary basis for exclusion alleged
- EX1022, EX1023, EX1025, EX1026:
 - Authenticated by EX1031
 - Not hearsay
 - Subject to hearsay exceptions under FRE 803(3), 803(6), and 807
- EX1002, paragraphs 62-67:
 - Based on personal knowledge
 - Corroborated by admissible exhibits
 - Supported by computer science expertise (EX1002, ¶66; EX1029, ¶3)