

Petitioner Microsoft's Demonstratives

Microsoft Corporation v. RealD Spark LLC

IPR2023-01046

U.S. Patent No. 10,740,985



I. Overview of the '985 Patent

II. Ground 1: Ganin

III. Ground 3: Chalom + Aslan

IV. Conclusion

I. Overview of the '985 Patent

'985 Patent: “Adjusting a Digital Representation of a Head Region”

(12) **United States Patent**
Sommerlade et al.

(10) **Patent No.:** **US 10,740,985 B2**
(45) **Date of Patent:** **Aug. 11, 2020**

(54) **ADJUSTING A DIGITAL REPRESENTATION OF A HEAD REGION**

(71) Applicant: **RealD Spark, LLC**, Beverly Hills, CA (US)

(72) Inventors: **Eric Sommerlade**, Oxford (GB);
Alexandros Neophytou, Oxford (GB)

(73) Assignee: **RealD Spark, LLC**, Beverly Hills, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/057,566**

(22) Filed: **Aug. 7, 2018**

(65) **Prior Publication Data**

US 2019/0051057 A1 Feb. 14, 2019

Related U.S. Application Data

(60) Provisional application No. 62/542,661, filed on Aug. 8, 2017, provisional application No. 62/543,587, filed on Aug. 10, 2017.

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,128,979	A	2/1915	Hess
1,970,311	A	8/1934	Ives
2,133,121	A	10/1938	Stearns

(Continued)

FOREIGN PATENT DOCUMENTS

CN	1142869	A	2/1997
CN	1377453	A	10/2002

(Continued)

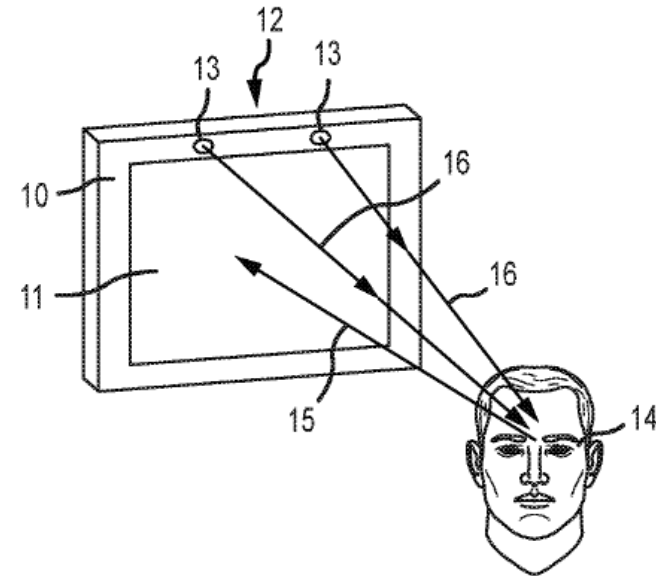
OTHER PUBLICATIONS

3M-™ ePrivacy Filter software professional version; http://www.cdw.com/shop/products/3M-ePrivacy-Filter-software-professional-version/3239412.aspx?cm_mmc=ShoppingFeeds-_-ChannelIntel-ligence-_-Software-_-3239412_3MT%20ePrivacy%20Filter%20software%20professional%20version_3MF-EPFPRO&cpncode=37-7582919&srcode=cii_10191459#PO; Copyright 2007-2016.

(Continued)

Primary Examiner — Jason A Pringle-Parker

(74) *Attorney, Agent, or Firm* — Penny L. Lowry; Neil G. J. Mothew



Ex. 1001, Fig. 1

The application relates particularly to adjusting target features of the digital representation of the head region, for example **to correct a perceived gaze direction of eyes**, or to modify the texture and/or shape of features such as the nose, mouth, chin or neck.

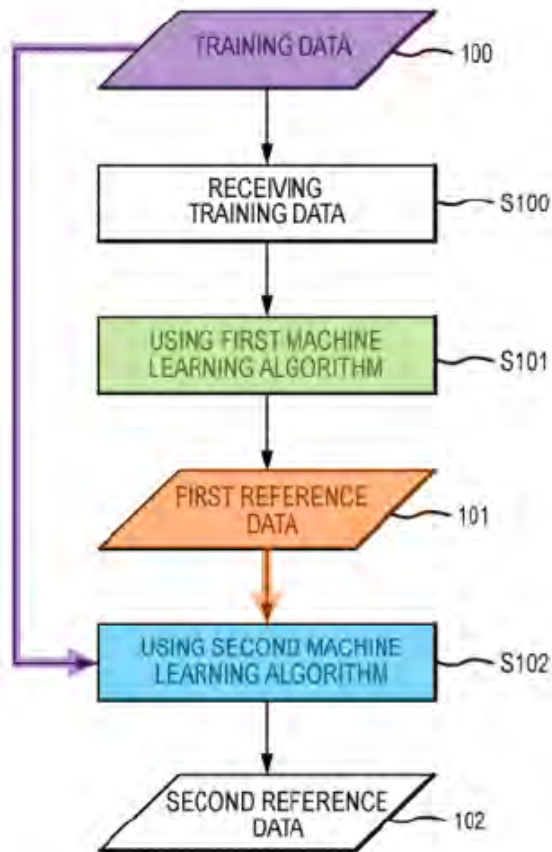
Ex. 1001, 1:9-12

'985 Patent: Challenged Claims

The Challenged Claims relate to eye-gaze correction, including training machine-learning models using pairs of input images (e.g., before and after gaze correction) and using the output of a first machine-learning algorithm (“MLA”) to train a second MLA

Paper 1 (Pet.) at 5

No.	'985 Patent Claim Elements
1[P]	A method of generating reference data for adjusting a digital representation of a head region, the method comprising:
1[A]	receiving training data comprising:
1[B]	a set of input patches, each input patch comprising a target feature of a digital representation of a head region prior to adjustment of the digital representation of the head region, wherein the target feature is the same for each input patch; and
1[C]	a set of output patches in one-to-one correspondence with the input patches, each output patch comprising the target feature of the digital representation of the head region after adjustment of the digital representation of the head region;
1[D]	using a first machine learning algorithm to generate first reference data using the training data , the first reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region; and
1[E]	using a second machine learning algorithm to generate second reference data using the same training data as the first machine learning algorithm and the first reference data generated by the first machine learning algorithm, the second reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region,
1[F]	wherein the first reference data comprise first editing instructions for a range of possible configurations of the target feature and first selection instructions for selecting editing instructions for a particular input patch from the first editing instructions based on the configuration of the target feature of the input patch; and
1[G]	wherein the second reference data comprise second editing instructions for a range of possible configurations of the target feature and second selection instructions for selecting editing instructions for a particular input patch from the second image editing instructions based on the configuration of the target feature of the input patch.



'985 Patent: Challenged Claims

PO does not dispute that Ganin or Chalom-Aslan discloses the claimed “training data” including “input patches” and “output patches”

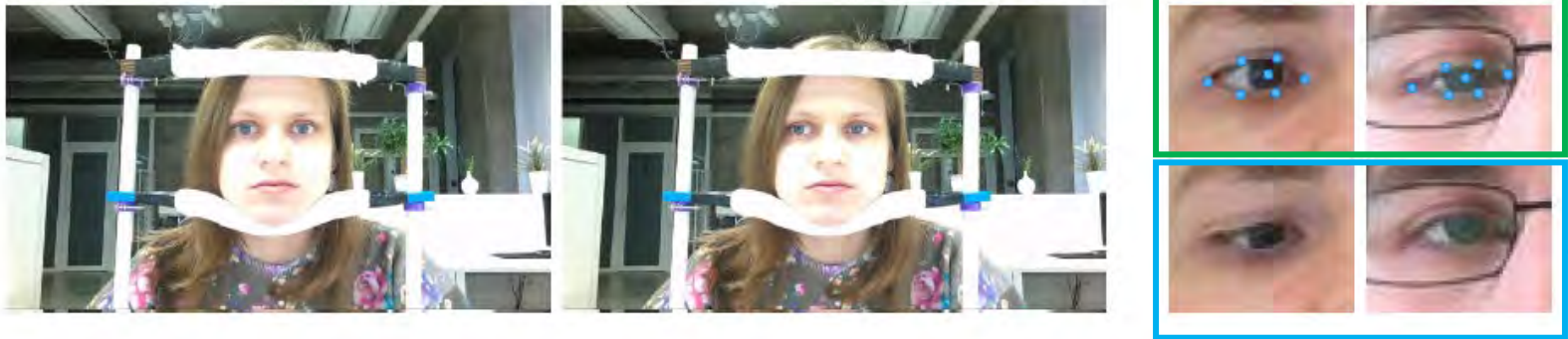


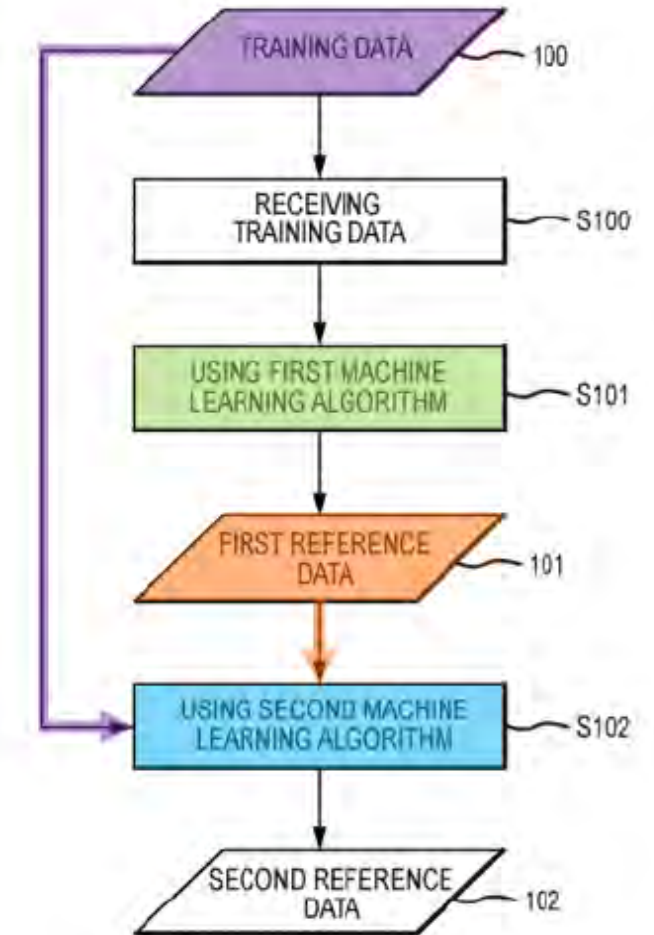
Fig. 4. Left – dataset collection process. Right – examples of two training pairs (input image with superimposed feature points on top, output image in the bottom).

Ex. 1006, Fig. 4

No.	'985 Patent Claim Elements
1[P]	A method of generating reference data for adjusting a digital representation of a head region, the method comprising:
1[A]	receiving training data comprising:
1[B]	a set of input patches , each input patch comprising a target feature of a digital representation of a head region prior to adjustment of the digital representation of the head region, wherein the target feature is the same for each input patch; and
1[C]	a set of output patches in one-to-one correspondence with the input patches, each output patch comprising the target feature of the digital representation of the head region after adjustment of the digital representation of the head region;

'985 Patent: Challenged Claims

No.	'985 Patent Claim Elements
1[D]	using a first machine learning algorithm to generate first reference data using the training data , the first reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region; and
1[E]	using a second machine learning algorithm to generate second reference data using the same training data as the first machine learning algorithm and the first reference data generated by the first machine learning algorithm, the second reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region,
1[F]	wherein the first reference data comprise first editing instructions for a range of possible configurations of the target feature and first selection instructions for selecting editing instructions for a particular input patch from the first editing instructions based on the configuration of the target feature of the input patch; and
1[G]	wherein the second reference data comprise second editing instructions for a range of possible configurations of the target feature and second selection instructions for selecting editing instructions for a particular input patch from the second image editing instructions based on the configuration of the target feature of the input patch.



'985 Patent: Claimed Elements all Well-Known



'985 Patent

The first machine learning algorithm [may comprise] a neural network [... or] a support vector machine or a generative adversarial network (GAN).

Ex. 1001, 16:24-37

The second machine learning algorithm may [comprise] a regression forest [... or] regression ferns, cluster centres, a lookup table, or separable filter banks [or] a second neural network.

Ex. 1001, 16:24-44

Reference to editing instructions herein is understood to encompass **any** data which can be used to define how a digital representation (e.g. image) of a head region should be adjusted to achieve a desired aim.

Ex. 1001, 18:21-24



Patent Owner's Expert
Eli Saber

“Claim 1 states you’re using the same training data. So you have the same training data. Now, how you use the data in your algorithms might differ slightly.”

Ex. 1061, 164:4-10

'985 Patent uses well-known “teacher-student” architecture



'985 Patent

“The second machine learning algorithm MLA2 thus gets the editing instructions to match or to approximate, and **does not have to infer these from the matched input images as MLA1.**”

Ex. 1001, 17:14-17



Petitioner's Expert
Lina Karam

“[A]lthough the '985 Patent does not use the term explicitly, it essentially adopts the well-known ‘teacher-student’ architecture for training a well-known machine learning based gaze correction algorithm.”

Ex. 1003, ¶¶67

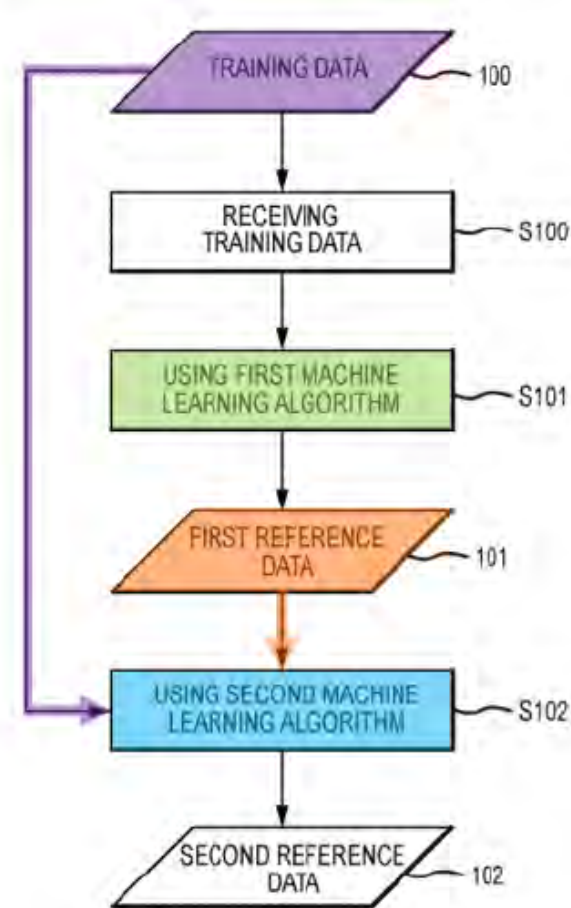


FIG. 8

Ex. 1001, Fig. 8

II. Ground 1: Ganin

Ganin Reference

DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation

Yaroslav Ganin, Daniil Kononenko, Diana Sungatullina, Victor Lempitsky

Skolkovo Institute of Science and Technology,
{ganin,daniil.kononenko,d.sungatullina,lempitsky}@skoltech.ru

Abstract. In this work, we consider the task of generating highly-realistic images of a given face with a redirected gaze. We treat this problem as a specific instance of conditional image generation and suggest a new deep architecture that can handle this task very well as revealed by numerical comparison with prior art and a user study. Our deep architecture performs coarse-to-fine warping with an additional intensity correction of individual pixels. All these operations are performed in a feed-forward manner, and the parameters associated with different operations are learned jointly in the end-to-end fashion. After learning, the resulting neural network can synthesize images with manipulated gaze, while the redirection angle can be selected arbitrarily from a certain range and provided as an input to the network.

Keywords: gaze correction, warping, spatial transformers, deep learning

1 Introduction

In this work, we consider the task of learning deep architectures that can transform input images into new images in a certain way (deep image resynthesis). Generally, using deep architectures for image generation has become a very active topic of research. While a lot of very interesting results have been reported over recent years and even months, achieving photo-realism beyond the task of synthesizing small patches has proven hard.

DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation

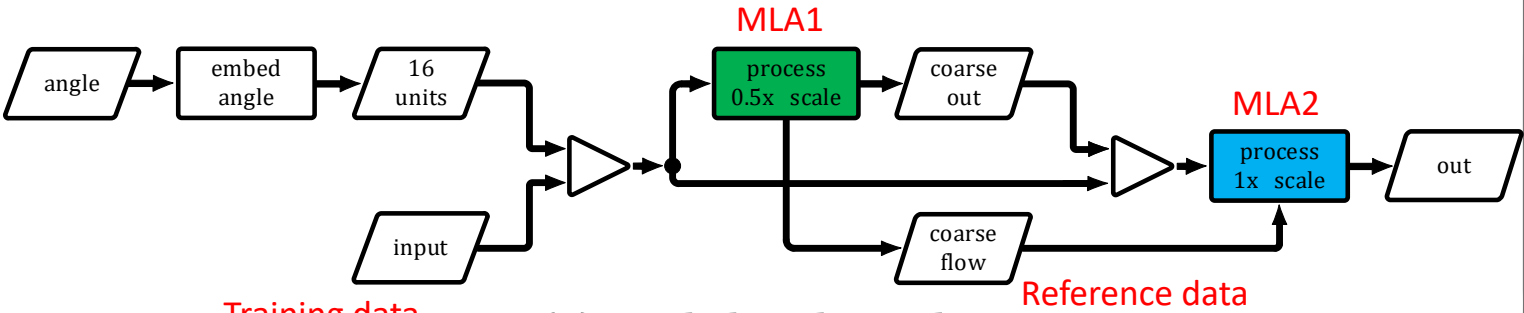
Yaroslav Ganin , Daniil Kononenko, Diana Sungatullina & Victor Lempitsky

Conference paper | [First Online: 17 September 2016](#)

18k Accesses | 50 Citations | 9 Altmetric

Part of the [Lecture Notes in Computer Science](#) book series (LNIP,volume 9906)

Ex 1008 at 1



(a) High-level pipeline

Ganin Reference



Petitioner's Expert
Lina Karam

“Ganin concatenates the input eye region, anchor points extracted from the eye region, and the desired angle into maps, which are then provided to the 2-scale flow NN that predicts a “warping field” to correct the gaze of an input image.”

“Ganin’s “multi-scale neural network” includes a coarse warping module and a fine warping module, both of which include a neural network.”

“The coarse warping module generates the coarse warping field and the fine warping module generates the residual warping field. The coarse warping field and the residual warping field are then combined together to warp the input eye region.”

Ex 1003 at ¶ 75

Ganin's fine warping module receives two separate inputs



Board

“Petitioner appears to rely on the forwarded reference data to teach the two separate inputs for the fine warping module”

Paper 7 at 43



Petitioner's Expert
Lina Karam

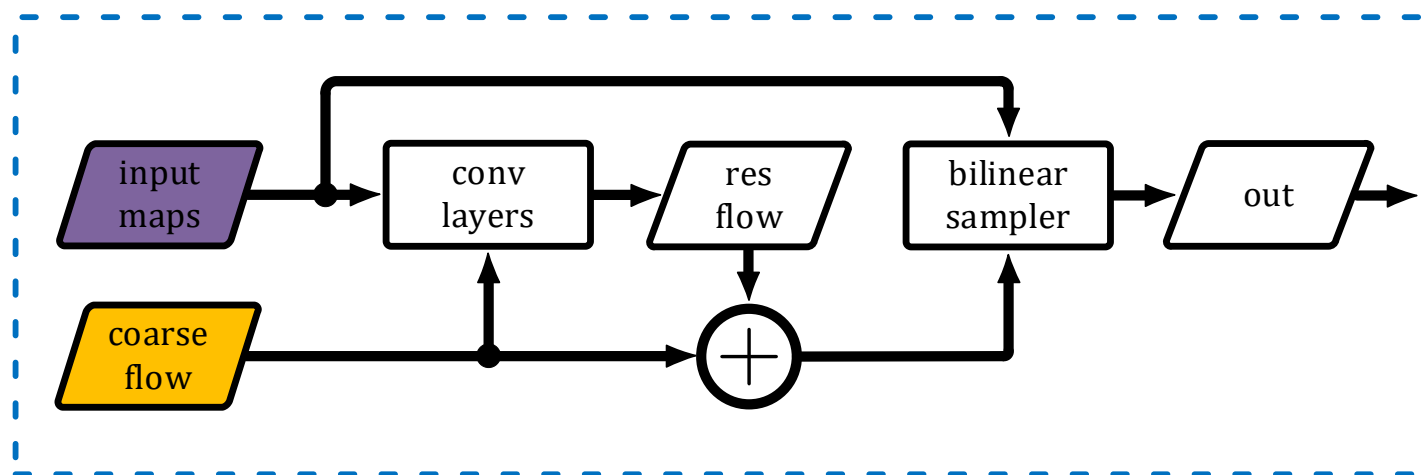
“Ganin’s fine warping module takes two inputs during runtime. First is the input maps. [] The other input to Ganin’s 1x-scale NN is the coarse warping field $D_{\text{coarse}}(I, \alpha)$ generated by the coarse warping module.”

Ex 1003 at 125

Ganin's fine warping module receives two separate inputs



Ganin



(c) 1x-scale processing module



Patent Owner's Expert
Eli Saber

“So you look [at Ganin Figure 10(c)], you see that’s essentially a blowup of the “process 1x scale.”

“And it receives as input [] the output from the concatenation. So that’s essentially your input maps. And it also receives as input the coarse flow that you see in the coarse flow box there. So you see two boxes going in that one [] You have the input maps and you have the coarse flow.”

Ex 1061 at 175:17 – 177:9

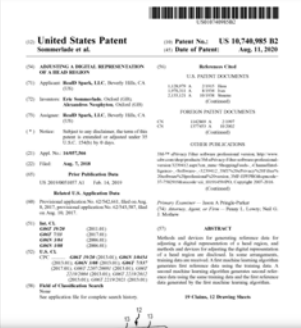
PO's Anti-Ganin Arguments Fail



PO argues that Ganin does not anticipate elements 1[D]/18[D] and 1[E]/18[E] of the Challenged Claims because:

- a) Ganin has 1 set of training data
- b) Ganin has 1 objective function
- c) Ganin's text uses singular forms

Paper 12 (POR) at 9-18



'985 Patent

No.	Claim Elements
1[D]	using a first machine learning algorithm to generate first reference data using the training data , the first reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region; and
1[E]	using a second machine learning algorithm to generate second reference data using the same training data as the first machine learning algorithm and the first reference data generated by the first machine learning algorithm, the second reference data comprising editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region,

Objective Function?

Ganin discloses two MLAs

1

Each MLA does NOT need its own training data

2

One objective function does NOT mean one MLA

3

Ganin's text uses both singular and plural forms

1

Each MLA does NOT need its own training data set

REAL D
Patent Owner

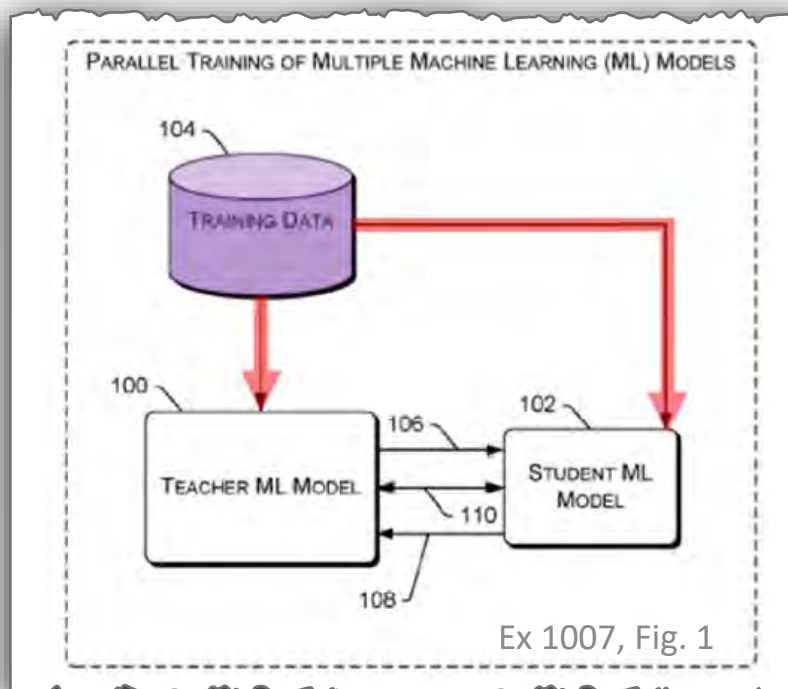
Different tasks?

“If Ganin’s coarse and fine modules were separate MLAs as Petitioner suggests, then each of these alleged MLAs would require different training data because these MLAs perform different tasks”

Paper 21 at 3-4



Aslan



Two Separate MLAs may use the same training data as is described in Aslan and the '985 Patent

Ganin's MLAs perform the same task: gaze redirection

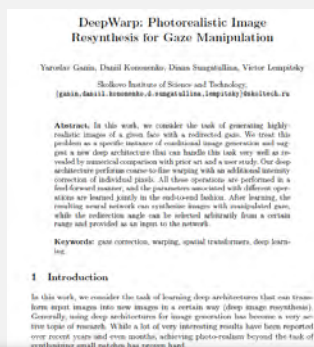


Petitioner's Expert
Lina Karam

Petitioner does not contend that Ganin requires two different tasks as alleged by PO

“During training, the NN can learn the corresponding warping-field generating function that covers all possible warping fields that can be generated for the considered task”

Ex 1003 at ¶ 134



Ganin

“In this work, we consider the task of generating highly-realistic images of a given face with a redirected gaze”

Ex 1006 at 1

“We concentrated on gaze redirection, although our approach might be extended for other similar tasks”

Ex 1006 at 13

“The model contains [...] two warping modules [...] predicting and applying pixel-flow to the input image”

Ex 1006 at 18

REAL D

Patent Owner

Patent Owner argues that Ganin does not disclose a “first machine algorithm . . . using the training data” and a “second machine algorithm . . . using the same training data as the first machine learning algorithm.” (claim elements 1[D]/[E] and 18[D]/[E])

Paper 12 at 18

Patent Owner argues that Ganin having 2 MLAs requires that the training data include coarse input/output training pairs and fine input/output training pairs.

Paper 12 at 18-19

Disputed claim elements recite simply “using” the training data



*Patent Owner's Expert
Eli Saber*

“Claim 1 states you’re using the same training data. So you have the same training data. Now, how you use the data in your algorithms might differ slightly.”

Ex. 1061, 164:4-10

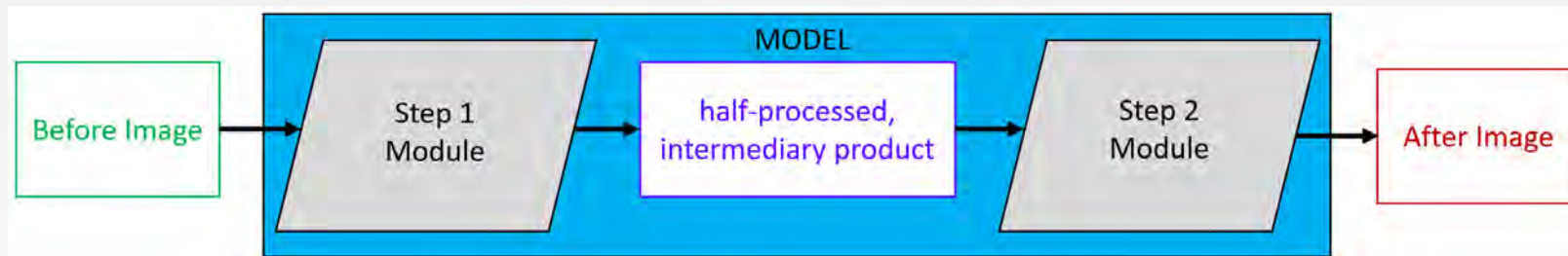
“Those (i.e., 1[E] and 18[E]) are the ones I opined on. I wasn’t asked to opine on any other elements.

Ex. 1061, 153:3-18

REAL D
Patent Owner

Patent Owner imagines a model with a “half-processed intermediary product” and argues that Ganin does not disclose such a product.

Paper 12 at 15



DeepWarp: Photorealistic Image
Resynthesis for Gaze Manipulation
Yaroslav Ganin, Dmitri Kononenko, Denis Sengulidze, Victor Lempitsky
[ganin, dmitri.kononenko, d.sengulidze, lempitsky@phystech.ru]

Abstract: In this work, we consider the task of generating highly-
realistic images of a given face with a manipulated gaze. We treat this
problem as a specific instance of conditional image generation and sug-
gest a new deep architecture that can handle this task very well as re-
sulted by numerical comparison with prior art and a user study. Our deep
architecture performs coarse-to-fine warping with an additional iterative
correction of individual pixels. All these operations are performed in a
feed-forward manner, and the parameters are used with different oper-
ations are learned jointly in the end-to-end fashion. After learning, the
resulting neural network can synthesize images with manipulated gaze,
while the underlying scale can be selected arbitrarily from a certain
range and provided as an input to the network.

Keywords: gaze correction, warping, spatial transformers, deep learn-
ing.

1 Introduction

In this work, we consider the task of learning deep architectures that can trans-
form input images into new images in a certain way (deep image resynthesis).
Generally, using deep architectures for image generation has become a very ac-
tive topic of research. While a lot of very interesting results have been reported
over recent years and even months, achieving photorealism beyond the task of
resynthesizing small patches has proven hard.

Ganin

Ganin discloses that each warping module (fine and coarse) produces a pixel-flow field (a two-channel map).

Ex 1006 at 5-6

Ganin discloses two MLAs



2

One objective function does NOT mean one MLA

REAL D
Patent Owner

“Ganin provides one objective function that assesses one output (i.e., redirected gaze image), and thus uses only one machine learning algorithm”

Paper 12 at 13

Neither Expert provided an MLA definition that requires individual objective functions or training sets



Patent Owner's Expert
Eli Saber

A machine learning algorithm is a computer model that trains itself to perform a specific task by using training data

Ex 2008 at ¶42



Petitioner's Expert
Lina Karam

A machine learning (“ML”) algorithm [] is an algorithm [] that learns from data to perform a task or a class of tasks

Ex 1003 at ¶23



When briefing the meaning of MLA in District Court, PO did not try to add any of the extra requirements it imagines here

Plaintiff's Proposed Construction for "machine learning algorithm":
"an algorithm that is able to learn from data"

Ex 1055 at 6



PO attempts to read limitations into the claims that are not even found in the written description

One of the cardinal sins of patent law – reading a limitation from the written description into the claims.

Phillips v. AWH., 415 F.3d 1303, 1320 (Fed. Cir. 2005)



*Patent Owner's Expert
Eli Saber*

Q: Is this something that a -- a skilled person in this field could have done, make use of one objective function to train multiple algorithms?

A: *You could.* I mean, it depends it -- it depends a lot -- I mean, you're -- you're asking me a question -- a very, very broad question.

Ex 1061 at 14:8-17

The '985 Patent does not mention
objective functions at all

Q: So looking at the '985 Patent, do you recall if it discusses an objective function?

A: I don't remember if the word "objective function" appear in the patent.

Q: Does Claim 1 reference an objective function?

A: Well, the words "objective functions" do not -- are not printed on that page literally, but obviously when you're doing the training, you have to utilize some type -- you -- you need to utilize some type of objective function to accomplish the training.

Ex 1061 at 117:13 – 118:1



Aslan

[0007] In some implementations, a process can include generating an objective function that is to be used for jointly training a set of machine learning models. The objective function can include at least one term that is a function of: (i) a first output of a first machine learning model and (ii) a second output of a second machine learning model. The process can further include optimizing the objective function to train the first machine learning model and the second machine learning model in parallel. In some implementations, optimizing the objective function includes determining values of model parameters, such as weight parameters, that optimize the objective function.

Ex 1007 at ¶17



Patent Owner's Expert
Eli Saber

Q: And does Aslan indicate whether one objective function can be used to train multiple machine-learning models?

A: Well, so if you look at paragraph thirty-five of Aslan and equation two that follows that paragraph, Aslan describes the objective function for joint training the first and second models.

Ex 1062 at 218:11-22

Ganin discloses two MLAs

3

Ganin's text uses both singular and plural forms

Ganin's text uses both singular and plural forms

REAL D
Patent Owner

“Since Ganin’s technical discussion relates to one algorithm, it should be unsurprising that Ganin repeatedly refers to its algorithm in the singular”

Paper 12 at 14

PO implies that the terms “neural network,” “model,” and “architecture” constitute an MLA but the term “module” does not.

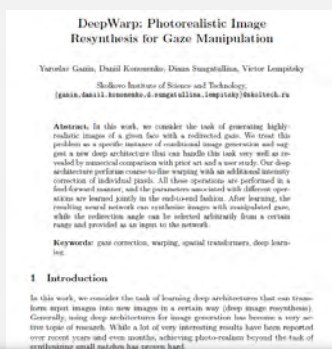
However, Ganin uses the terms interchangeably

The output is enhanced by processing with the lightness correction neural network.

Ex 1006 at 4

Figure 11 shows the actual architecture for the lightness correction module.

Ex 1006 at 17



Ganin

DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation

Yaroslav Ganin, Daniel Kozminski, Dima Sengupta, Victor Lempitsky
Skolov Institute of Science and Technology
{ganin, daniil.kozminski, d.sengupta, v.lempitsky}@skoltech.ru

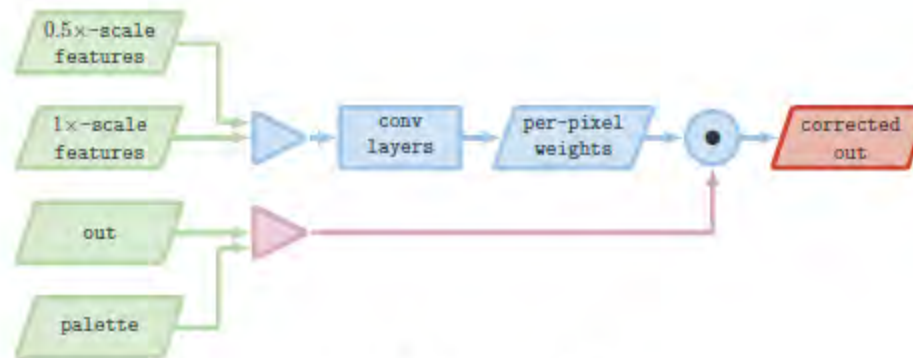
Abstract. In this work, we consider the task of generating highly-realistic images of a given face with a redirected gaze. We treat this problem as a specific instance of conditional image generation and suggest a new deep architecture that can handle this task very well as revealed by numerical comparison with prior art and a user study. Our deep architecture performs coarse-to-fine resynthesing with an additional intensity correction of individual pixels. All these operations are performed in a feed-forward manner, and the parameters associated with different operations are learned jointly in the end-to-end fashion. After learning, the resulting neural network can resynthesize images with manipulated gaze, while the selection angle can be selected arbitrarily from a certain range and provided as an input to the network.

Keywords: gaze correction, warping, spatial transformers, deep learning.

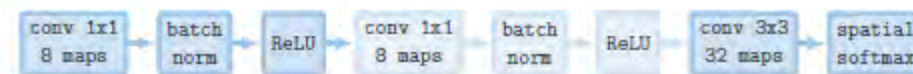
1 Introduction

In this work, we consider the task of learning deep architectures that can transform input images into new images in a certain way (deep image resynthesis). Generally, using deep architectures for image generation has become a very active topic of research. While a lot of very interesting results have been reported over recent years and even months, achieving photo-realism beyond the task of synthesizing small patches has proven hard.

Ganin



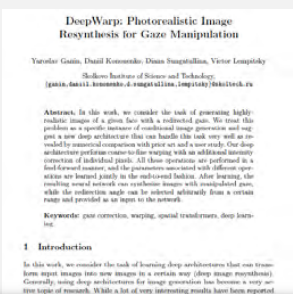
(a) Architecture.



(b) Convolutional layers.

Fig. 11. Lightness Correction Module increases lightness of selected regions. **11(a)** shows the actual architecture of the module. Multi-scale features are processed by the convolutional neural network presented in **11(b)**.

Ex 1006 at 18



Ganin

Appendix B.1 Warping stage

A flowchart of the pipeline *without* the lightness correction module is depicted in Figure 10. To allow for more interaction between the input data and the correction angle embedding, we choose not to perform late fusion [6] and feed the embedding vector as an additional input to the warping network. More concretely, we replicate the 16-dimensional $F_{\text{angle}}(\alpha)$ for every spatial location creating a tensor of size $16 \times H \times W$ which is then concatenated to the rest of the input (*pink* triangle in Figure 10(a)). The architectures of the two warping modules are same (modulo the number of input maps) and are shown in Figure 10(d).

Ex 1006 at 17



Patent Owner's Expert
Eli Saber

Q: So it -- does it not say here that an architecture, that single feed-forward architecture, can include multiple, plural, architectures? Is that not what it's saying?

A: Well, you're taking a little bit -- things a little bit out of context. So essentially -- **he could have used better wording**, but he's essentially referring -- if you look at the next figure -- the next page, if you look at Figure 10(d), those are the convolutional layers that he's utilizing in his high level pipeline and then in his B and C.

Q: Okay. So there are -- there are multiple architectures within that single feed-forward architecture; right?

A: Well, the -- **the choice of words is** -- is not -- is **not ideal** in the way it's stated here.



“As to Petitioner’s argument that the “lightness correction” module is a standalone MLA, this was not argued by Petitioner prior to its Reply and as such, should be disregarded on that basis alone”

Paper 21 at 8



A petitioner may file a reply to a patent owner response [...] to respond to arguments raised in the preceding brief. 37 C.F.R. § 42.23.

U.S. Patent Trial and Appeal Board, Consolidated Trial Practice Guide (Nov. 2019) at 73-74

Countering an argument in a prior brief is not a belated argument.

Idemitsu Kosan Co., LTD., v. SFC CO. LTD., No. 2016-2721 (Fed. Cir. Sept. 15, 2017)

PO never argued that Ganin lacked two MLAs during European Prosecution



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

The EPO determined that Ganin was novelty destroying for the claims of the corresponding EP Application, asserting that Ganin disclosed limitations 1[D] and 1[E].

Ex 1013 at 214, 226-227

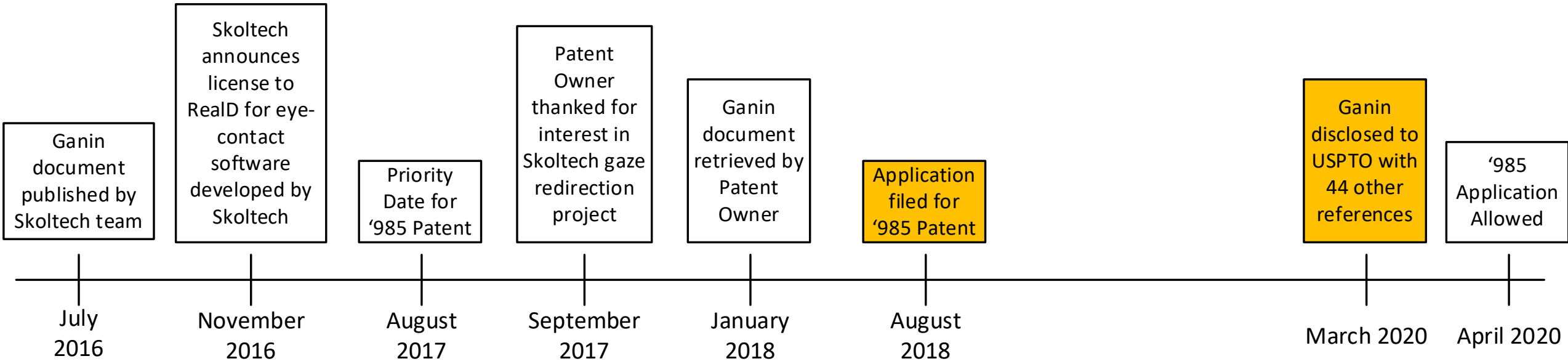
Patent Owner did not argue that Ganin lacked two MLAs.

Ex 1013 at 236-237

Patent Owner added limitations related to compressing editing instructions using a basis set, which are not present in any of the Challenged Claims.

Ex 1013 at 236-237

PO's Early Skoltech Knowledge and Belated Ganin Disclosure



III. Ground 3: Chalom + Aslan

Undisputed: '985 Patent adopts a “teacher-student” architecture



'985 Patent

“The second machine learning algorithm MLA2 thus gets the editing instructions to match or to approximate, and **does not have to infer these from the matched input images as MLA1.**”

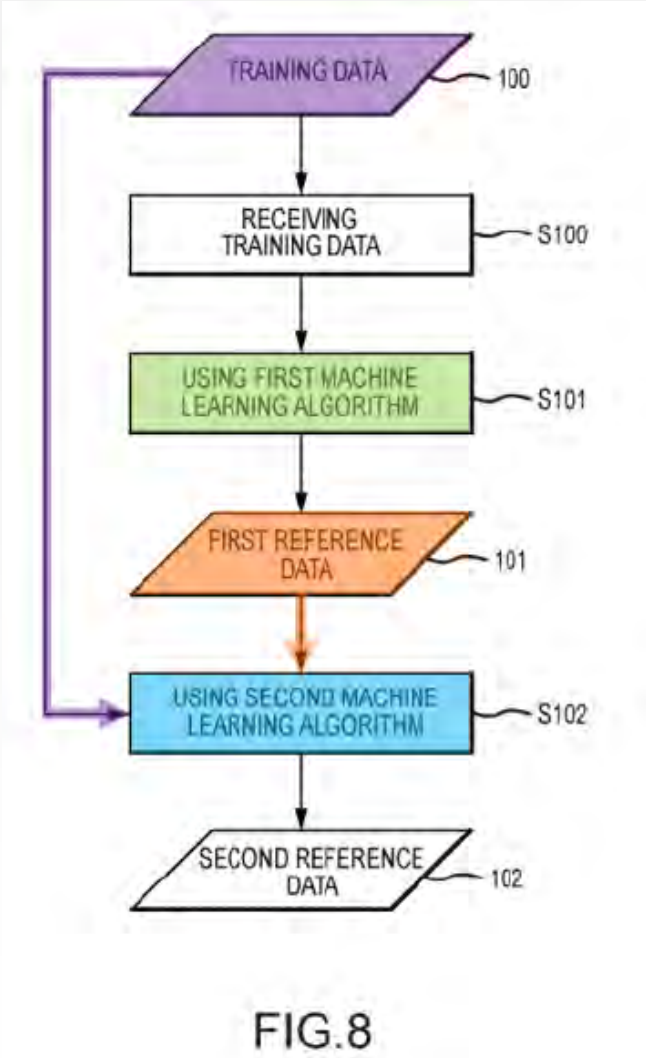
Ex. 1001, 17:14-17



Petitioner's Expert
Lina Karam

“[A]lthough the '985 Patent does not use the term explicitly, it essentially adopts the well-known ‘teacher-student’ architecture for training a well-known machine learning based gaze correction algorithm.”

Ex. 1003, ¶¶67



Ex. 1001, Fig. 8

Primary Reference: Chalom

(12) **United States Patent**
Chalom et al.

(10) **Patent No.:** **US 10,664,949 B2**
(45) **Date of Patent:** **May 26, 2020**

(54) **EYE CONTACT CORRECTION IN REAL TIME USING MACHINE LEARNING**

(71) Applicant: **Intel Corporation**, Santa Clara, CA (US)

(72) Inventors: **Edmond Chalom**, Petach Tikvah (IL);
Or Shimshi, Kiryat-Ono (IL)

(73) Assignee: **Intel Corporation**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 885 days.

(21) Appl. No.: **15/136,621**

(22) Filed: **Apr. 22, 2016**

(65) **Prior Publication Data**
US 2017/0308992 A1 Oct. 26, 2017

(56) **References Cited**
U.S. PATENT DOCUMENTS
6,714,234 B1 3/2004 Hillis et al.
2003/0197779 A1* 10/2003 Zhang H04N 7/144
2005/0089246 A1 4/2005
2009/0238419 A1 9/2009
2012/0219180 A1 8/2012
(Co
FOREIGN PATENT DOCUMENTS
KR 1020140076949
KR 1020150113795
(Co
OTHER PUBLICATIONS
Song, F., Tan, X., Liu, X., Chen, Y., et al. "Detecting eye contact in still images with multi-scale hierarchical features." Pattern Recogn. 47(9), 1685-1698 (2013).

```
graph LR
    subgraph 800
        direction LR
        SI[SI 812] --> 801[Face Detection & Facial Landmark Detection 801]
        801 -- SFL 814 --> 803[Crop & Resize 803]
        803 -- SER 816 --> 805[Features Determination 805]
        805 -- SFS 818 --> 807[Random Forest Generator 807]
        TI[TI 813] --> 802[Face Detection & Facial Landmark Detection 802]
        802 -- TFL 815 --> 804[Crop & Resize 804]
        804 -- TER 817 --> 806[Motion Vector Candidates 806]
        806 -- LM 819 --> 807
        807 -- RFC 820 --> Out[ ]
    end
```

FIG. 8

Ex. 1005, Fig. 8

Microsoft's Demonstratives / Not Evidence

35

Secondary Reference: Aslan

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2017/0132528 A1**
Aslan et al. (43) **Pub. Date: May 11, 2017**

(54) **JOINT MODEL TRAINING**

(71) Applicant: **Microsoft Technology Licensing, LLC**,
Redmond, WA (US)

(72) Inventors: **Ozlem Aslan**, Edmonton (CA); **Rich Caruana**, Woodinville, WA (US); **Matthew R. Richardson**, Seattle, WA (US); **Abdelrahman Mohamed**, Redmond, WA (US); **Matthai Philipose**, Redmond, WA (US); **Krzysztof Geras**, Edinburgh (GB); **Gregor Urban**, Munich (DE); **Shengjie Wang**, Bellevue, WA (US)

(21) Appl. No.: **15/195,894**

(22) Filed: **Jun. 28, 2016**

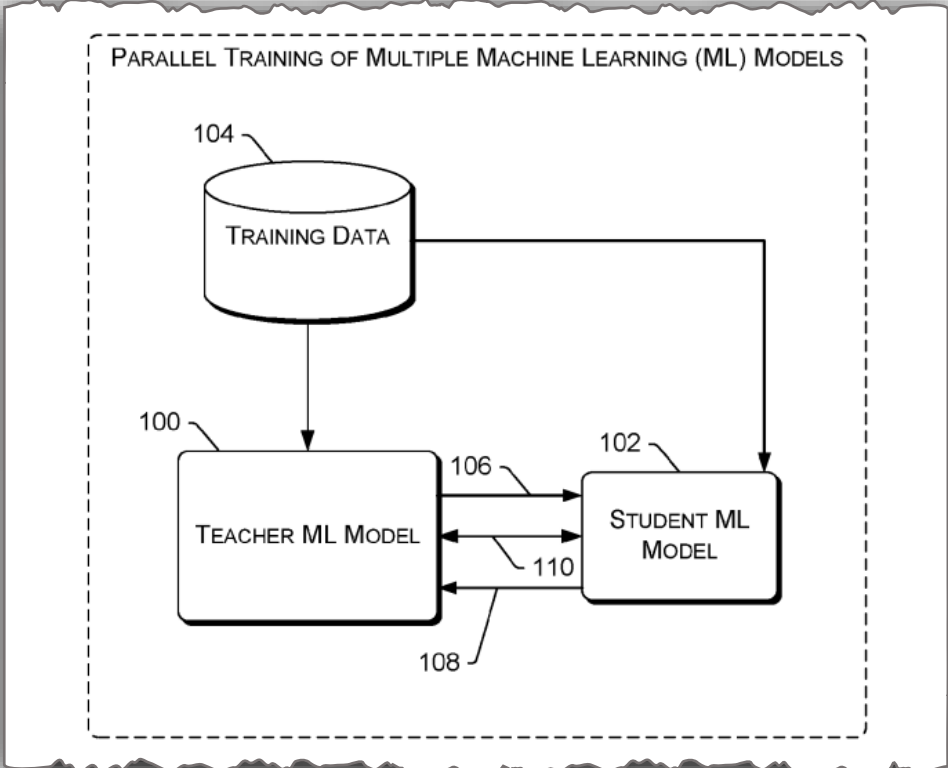
Related U.S. Application Data

(60) Provisional application No. 62/252,355, filed on Nov. 6, 2015.

Publication Classification

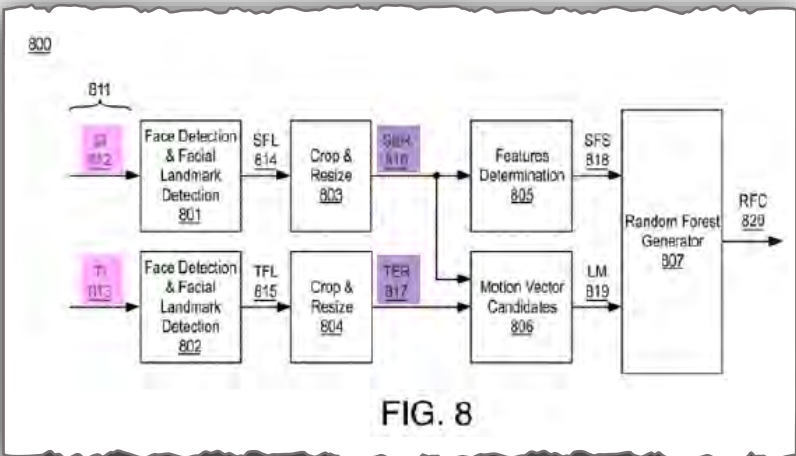
(51) **Int. Cl.**
G06N 99/00 (2006.01)
G06N 7/00 (2006.01)
(52) **U.S. Cl.**
CPC **G06N 99/005** (2013.01); **G06N 7/005** (2013.01)

(57) **ABSTRACT**
Multiple machine learning models can be jointly trained in parallel. An example process for jointly training multiple machine learning models includes providing a set of machine learning models that are to learn a respective task, the set of machine learning models including a first machine learning model and a second machine learning model. The process can initiate training of the first machine learning model to learn a task using training data. During the training of the first machine learning model, information can be passed between the first machine learning model and the second machine learning model. Such passing of information (or “transfer of knowledge”) between the machine learning models can be accomplished via the formulation, and optimization, of an objective function that comprises model parameters that are based on the multiple machine learning models in the set.



Chalom-Aslan Combination

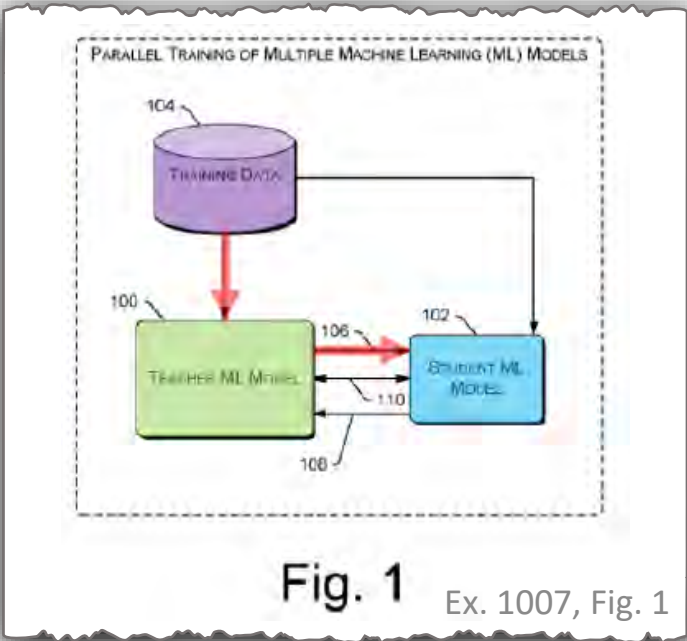
Ex. 1005, Fig. 8



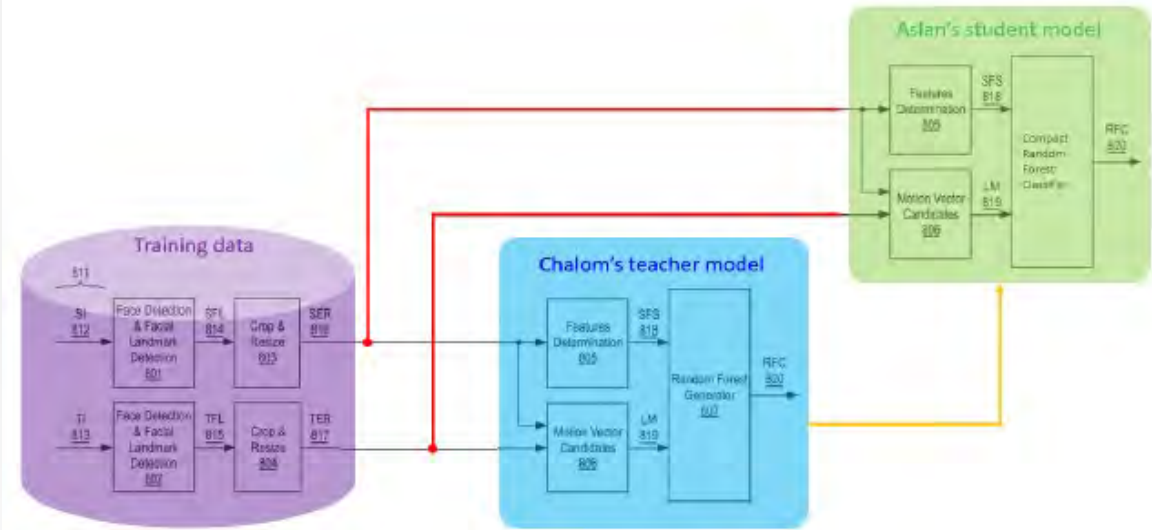
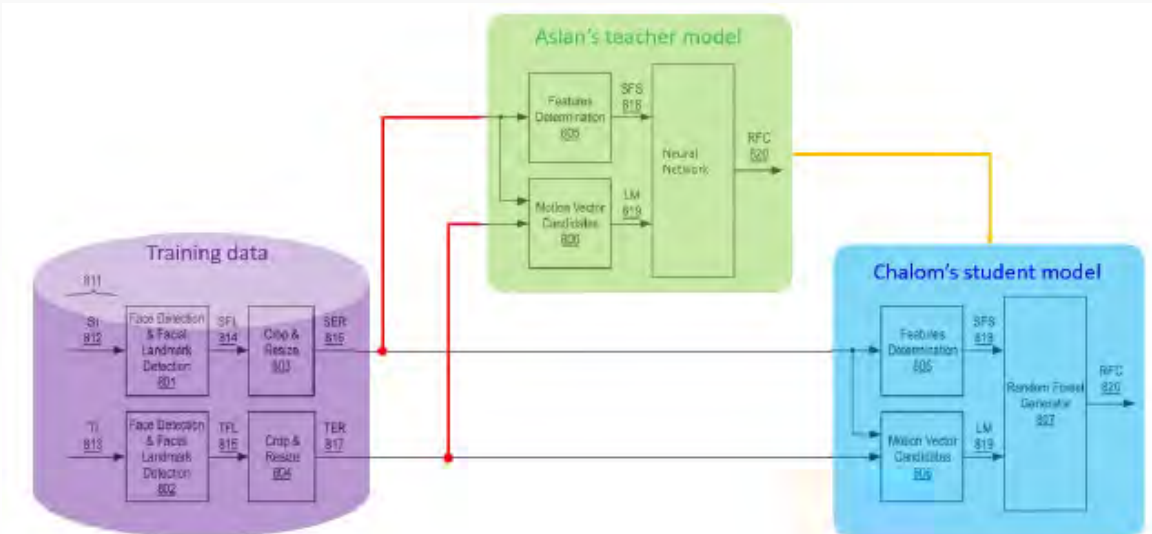
Chalom



Aslan

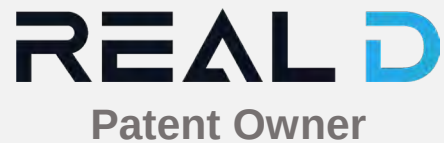


Ex. 1007, Fig. 1



III.A Chalom-Aslan teaches all claim elements

PO wrongly contends that Chalom-Aslan lacks “first reference data” in 1[E]



“[T]he second machine learning algorithm **must also** receive and use the first machine learning algorithm’s selection instructions, not only the editing instructions.”

POR, 28

“Chalom-Aslan does not disclose a second algorithm that receives both the first editing instructions and the first selection instructions as recited in ’985 claim 1.”

PO Sur-reply, 16

“first reference data” in 1[E]



'985 Patent

'985 Patent, Claim 1

1[D] using a first machine learning algorithm to generate first reference data using the training data, **the first reference data comprising editing instructions...**; and

1[E] using a second machine learning algorithm to generate second reference data **using** the same training data as the first machine learning algorithm and **the first reference data generated by the first machine learning algorithm...**

1[F] wherein **the first reference data** comprise **first editing instructions** for a range of possible configurations of the target feature **and first selection instructions** for selecting editing instructions for a particular input patch from the first editing instructions based on the configuration of the target feature of the input patch...

Ex. 1001, Claim 1

'985 Patent's MLA2 does not "receive and use" first selection instructions



'985 Patent

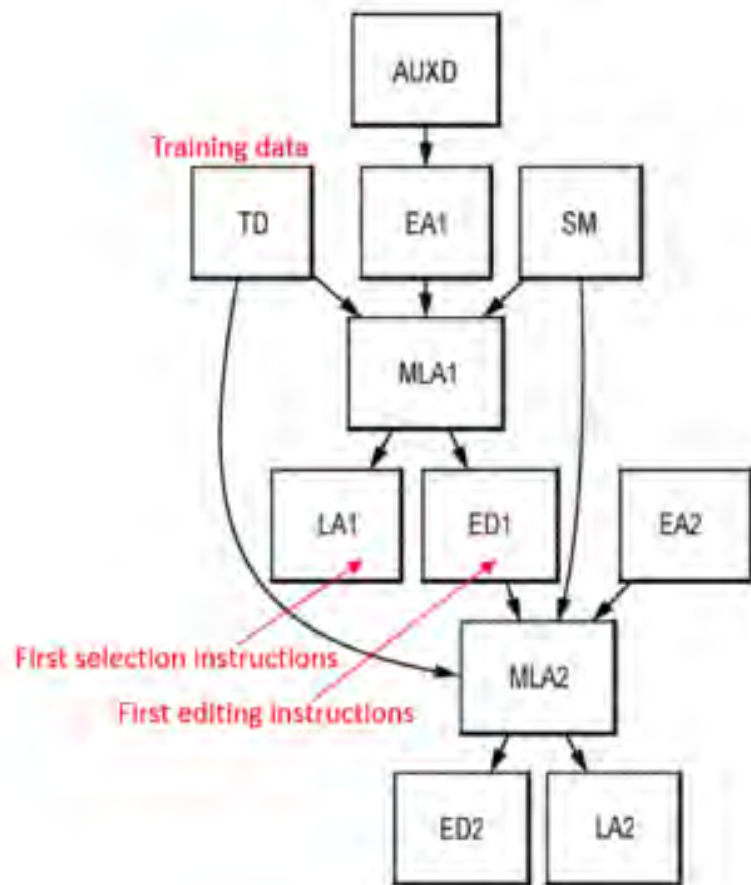


FIG. 9

Ex. 1001, Fig. 9

The first machine learning algorithm **MLA1** outputs **first selection instructions LA1** and **first editing instructions ED1**. The second machine learning algorithm **MLA2** receives the same training data **TD** and, optionally, the second editing algorithm **EA2**, and the similarity metric **SM**. **The second machine learning algorithm MLA2 additionally receives the first editing instructions ED1.**

Ex. 1001, 17:5-12

PO's expert agrees that '985 Patent's MLA2 only receives editing instructions ED1



Patent Owner's Expert
Eli Saber

1 Q. -- there's an arrow -- there's an arrow
2 going to MLA2, but if we look at LA1, we don't see
3 an arrow going to MLA2.
4 So I was just wondering, what's -- what's
5 the difference? Why does one have an arrow flowing
6 to MLA2 and the other does not? What does that
7 indicate?
8 A. Well, again, you're looking at a block
9 diagram, so it -- it is a high overview block
10 diagram.
11 So essentially what it's saying is -- and,
12 again, this is an overview -- the editing
13 instructions that MLA1 has generated are available
14 to MLA2 to utilize.
15 Q. And what about -- it doesn't look like LA1
16 has an arrow going to MLA2.
17 What's the significance of that?
18 A. Well, it's providing the editing
19 instruction to MLA2, and that's essentially part of
20 their process.

Ex. 1061, 145:1-20

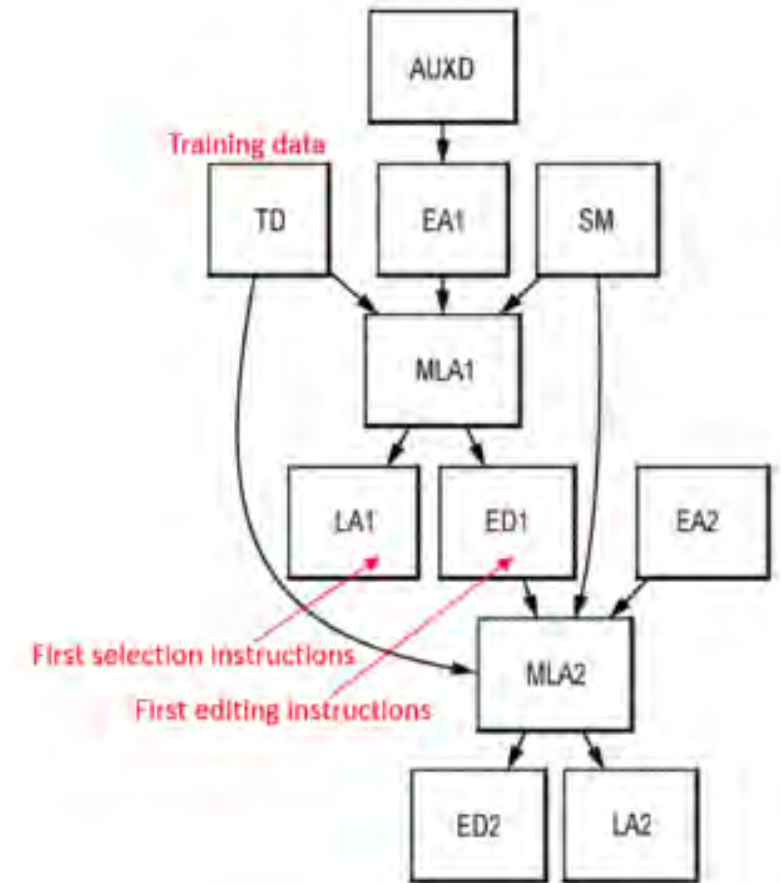


FIG.9

Ex. 1001, Fig. 9

Runtime v. Training: First reference data in Claim 1



'985 Patent

'985 Patent, Claim 1

1[D] using a first machine learning algorithm to generate first reference data using the training data, the **first reference data comprising editing instructions** for adjusting the digital representation of the head region for a range of possible digital representations of the head region; and

1[E] using a second machine learning algorithm to generate second reference data **using** the same training data as the first machine learning algorithm and **the first reference data generated by the first machine learning algorithm...**

1[F] wherein **the first reference data comprise first editing instructions** for a range of possible configurations of the target feature **and first selection instructions** for selecting editing instructions for a particular input patch from the first editing instructions based on the configuration of the target feature of the input patch...

Ex. 1001, Claim 1

Two different types of editing instructions are not passed to MLA2



'985 Patent

'985 Patent, Claim 1

1[D] using a first machine learning algorithm to generate first reference data using the training data, the **first reference data comprising editing instructions** for adjusting the digital representation of the head region for a range of possible digital representations of the head region; and

1[E] using a second machine learning algorithm to generate second reference data **using** the same training data as the first machine learning algorithm and **the first reference data generated by the first machine learning algorithm...**

1[F] wherein the **first reference data comprise first editing instructions** for a range of possible configurations of the target feature **and first selection instructions** for selecting editing instructions for a particular input patch from the first editing instructions based on the configuration of the target feature of the input patch...

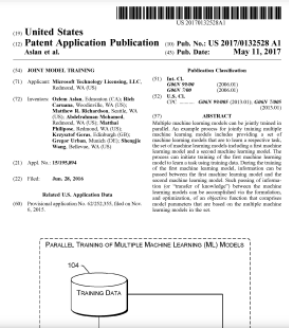
Ex. 1001, Claim 1

PO Misunderstands Chalom-Aslan: No Passing Objective Function



“When a POSITA does so, they do not arrive at a combination that practices the ’985 claims. Indeed, they arrive at a **teacher-student system that exchanges an ‘objective function,’** which cannot be the claimed ‘first reference data.’”

POR, 32



Aslan

“**passing information** between the models 100 and 102 via the objective function.”

Ex. 1007, ¶139

PO Misunderstands Chalom-Aslan: No Passing Objective Function



Petitioner's Expert
Lina Karam

“The teacher’s output, e.g., **first editing instructions including motion vector fields**, is used to train the student that also generates motion vector fields, i.e., second editing instructions.”

Ex. 1003, ¶274

“If you’re **passing information** from the teacher and you wanted to -- and you want to use this information to teach the student, you have to then **include it somehow in your optimization**. And, typically, **this is a term in your objective function**, so you take it, and you try to minimize the distance. That’s how you learn.”

Ex. 2009, 105:14-20

PO Misunderstands Chalom-Aslan: Different Reference Data



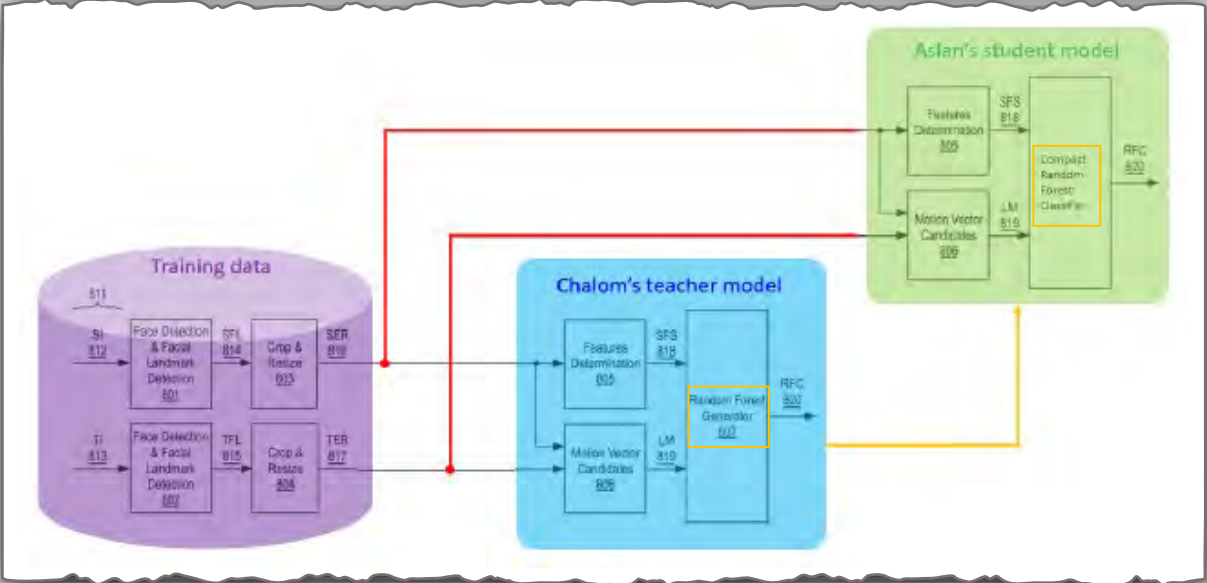
“Petitioner’s Chalom-Aslan combination fails to disclose the claimed ‘first reference data’ and ‘second reference data’... **because the Chalom-Aslan combination uses the same inputs, provided to the same modules, which would return the same outputs (i.e., the same “reference data”).”**

POR, 36

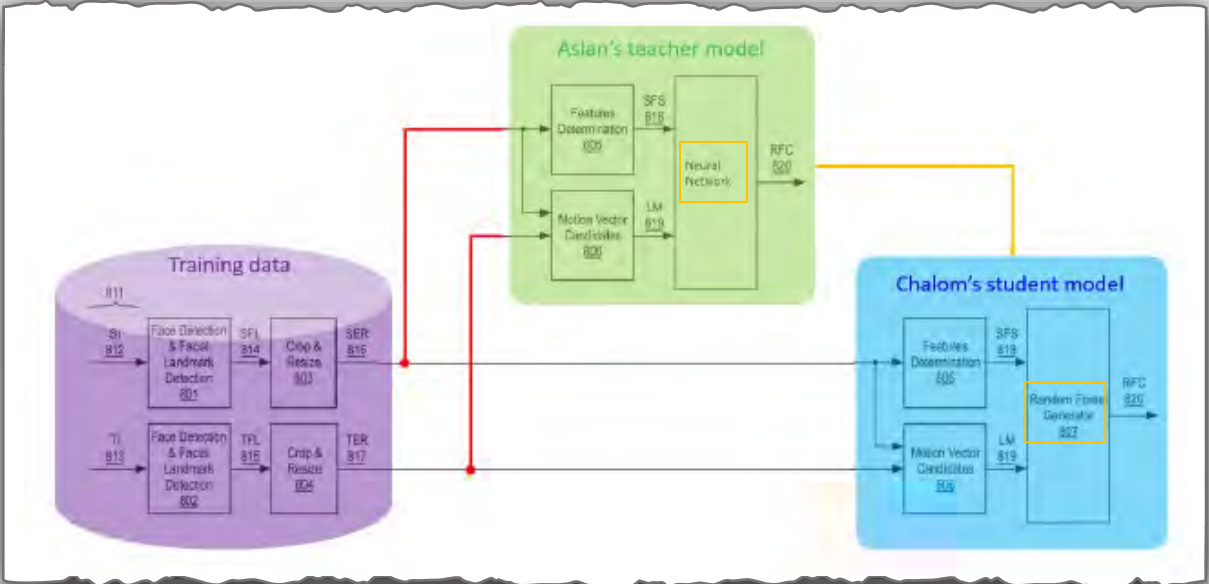
PO Misunderstands Chalom-Aslan: Different Reference Data



Petitioner's Expert
Lina Karam



Ex. 1005, Fig. 8, modified based on Aslan to incorporate a student model



Ex. 1005, Fig. 8, modified based on Aslan to add a teacher model

PO Misunderstands Chalom-Aslan: Different Reference Data



Board

“Petitioner then proposes modifying Chalom’s system with Aslan’s teacher-student MLAs to train the two MLAs **with the same input data** to produce a motion field output, as well as to provide the motion vector correction from the teacher MLA to the student MLA for further refinement.”

Institution Decision, 36

PO Misunderstands Chalom-Aslan: Different Modules



*Petitioner's Expert
Lina Karam*

“One way is using Chalom’s RFC as a teacher model to train **an even more compact MLA, e.g., a smaller random forest.**”

Ex. 1003, ¶219

“[A] PHOSITA would have been motivated to... us[e] Chalom’s already trained RFC as a teacher **to train a more compact RFC with fewer and/or less complex decision trees...**”

Ex. 1003, ¶222

“[A] PHOSITA... would have been motivated to... **train[] a new accurate teacher model** first, which is then used to train **Chalom’s RFC as a student model**. As taught by Aslan, **such new teacher model could be, e.g., a deep neural network or another type of machine learning model.**”

Ex. 1003, ¶224

III.B PO's Anti-Motivation Arguments Fail

PO wrongly argues lack of motivation to combine Chalom with Aslan



“[T]he Chalom-Aslan combination (Grounds III – V) does not render the ’985 Patent claims obvious **because there is no motivation—express or implied—to combine these references.**”

PO Sur-reply at 2

Motivations to combine Chalom with Aslan

1

Express motivation to combine

2

More Motivation: Known Elements and Accuracy and/or Efficiency Gains

3

Chalom-Aslan Combo is Sufficiently Explained

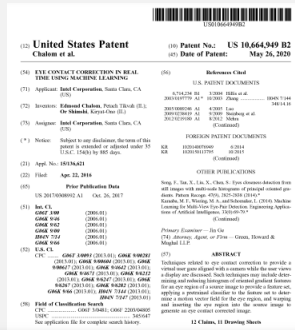
1 Express Motivation

“The pretrained classifier may include any suitable pretrained classifier such as a **random forest classifier** and the pretrained classifier **may be trained using *any suitable technique or techniques.***”

Ex. 1005, 14:62-67

“The pretrained classifier **may be compressed *using any suitable technique or techniques.*** In an embodiment, the pretrained classifier may be compressed based on **parameterized surface fitting.**”

Ex. 1005, 15:4-9



Chalom

1 Express Motivation



*Patent Owner's Expert
Eli Saber*

Q. And would **parameterized surface fitting** have been the **only example of a suitable compression technique** that a POSITA would have known... at this time period?

A.... **That is one suitable way** to do it.

Q. And what would have been some **other suitable ways**, considering that you're familiar with the Chalom reference...

A. ...you could also, for instance,... reduce a number of features... So **that's another form of compression** in a sense. **So it depends...**

Ex. 1061, 185:17-186:23

1 Express Motivation



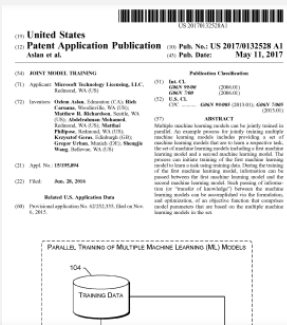
*Patent Owner's Expert
Eli Saber*

Q. And... where would the POSITA in this time period, pre-2017, go to look to figure out what that optimal compression would look like and what the suitable technique or techniques might have been?

A. Well, **they may look at some books**, they may... based on their experience and knowledge in the – in the field.

Ex. 1061, 187:15-24

1 Express Motivation



Aslan

“Often, the **best performing** machine learning models—in terms of the accuracy of the model’s output—comprise **ensembles** of hundreds or thousands of base-level machine learning models... [B]ecause ensembles typically require a relatively large storage footprint and powerful processing resources to execute at runtime, they are **not well suited for** implementations where storage space and/or computational power is at a premium (such as with **smart phones**, wearables, hearing aids, etc.).”

Ex. 1007, ¶13

1 Express Motivation



*Patent Owner's Expert
Eli Saber*

Q. [W]hy do **ensembles** typically require a typically **large storage footprint and powerful resources to execute** at runtime?

A. Well, if you have two different machine-learning architectures, each of them have 100 layers and each layer has a hundred neurons,... that is a **large architecture** that you have to train all of that... **[which] requires a lot of time and resources.**

Ex. 1062, 227:20-228:3

Q. And so could – a **random forest classifier**, could that be an **ensemble** in the pre 2017 time frame?

A. Generally speaking, **yes.**

Ex. 1062, 227:3-5

1 Express Motivation



Aslan

“Because the smaller second model 102 take much less memory to maintain and can operate faster on less processing power at runtime, **the second model 102 can be a compressed form of the larger first model 100** such that the second model 102 can be more readily deployed on computing devices with limited resources (e.g., mobile devices, wearables, etc.).”

Ex. 1007, ¶44



Patent Owner's Expert
Eli Saber

Q. So is Aslan telling a POSITA in the pre 2017 time frame that its joint training techniques can be used for model compression?

A. Yes. He is making that statement, as well as stating **one example application is model compression.**

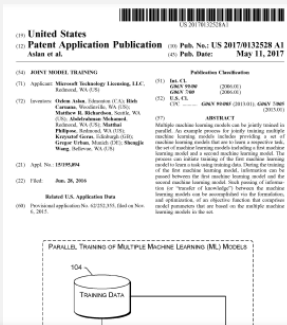
Ex. 1062, 239:16-22

Motivation: Known Elements Performing Ordinarily



2

More Motivation: Known elements performing ordinary functions



Aslan

“For example, **suitable machine learning models** for use with the techniques and systems described herein **include**, without limitation, tree-based models, support vector machines (SVMs), kernel methods, **neural networks**, **random forests**, splines (e.g., multivariate adaptive regression splines), hidden Markov model (HMMs), Kalman filters (or enhanced Kalman filters), Bayesian networks (or Bayesian belief networks), expectation maximization, genetic algorithms, linear regression algorithms, nonlinear regression algorithms, logistic regression-based classification models, or an ensemble thereof.”

Ex. 1007, ¶22



Patent Owner's Expert
Eli Saber

Q. And so **these types of machine-learning models** that Aslan lists in paragraph twenty-two, those were **known elements to a POSITA in the pre 2017 time frame?**

A. Yes. They would be known to a POSITA.

Ex. 1062, 216:6-217:3



*Patent Owner's Expert
Eli Saber*

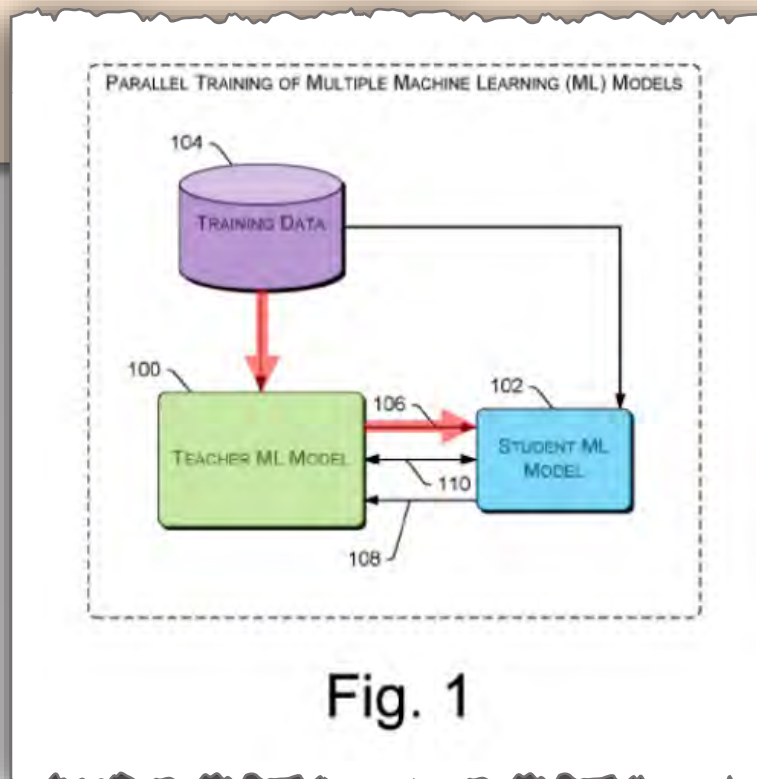
Q. Could a **neural network** have been the **machine-learning model 100**, consistent with the **Aslan** reference?

A. **That is what it states.**

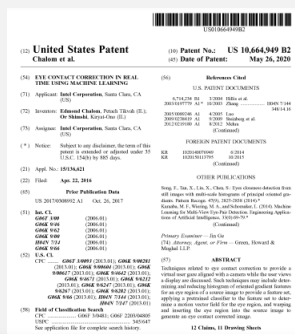
Q. ... Could [the **machine-learning model 102**] been a **random forest** consistent with the Aslan reference?

A. **Again, that is what it states...**

Ex. 1062, 220:23-221:12



Ex. 1007 (Aslan), Fig. 1



Chalom

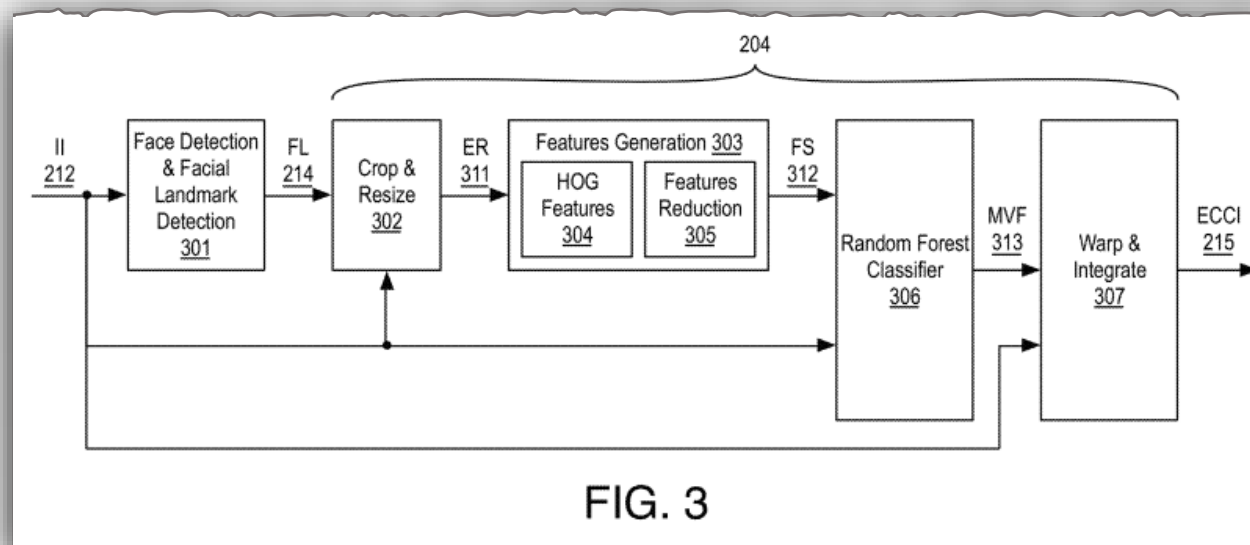


FIG. 3

Ex. 1005 (Chalom), Fig. 5



**Patent Owner's Expert
Eli Saber**

A. ... And then you have a feature generation, Section 303, that does the feature generation. And then **the random forest classifier**, we use the feature sets, that FS 312, together with the original image, it **will do the classification**.

Ex. 1061, 135:18-136:3



*Patent Owner's Expert
Eli Saber*

Q. ... [W]hat would be the **ordinary function of a random forest classifier** in the pre 2017 time frame?

A. The function of a random forest classifier is **to perform classification...** [F]or instance,... **figure three of Chalom, you can see a random forest classifier labeled 306** who takes in as input the feature sets from the previous block, 303, and **performs a classification and provides a motion vector field.**

Ex. 1062, 230:19-231:16



*Patent Owner's Expert
Eli Saber*

Q. In that last sentence of paragraph two of Exhibit 1007, does **Aslan** provide an example of **a task that can be performed with the system that Aslan describes**?

... ..

A. Yeah. So Aslan states: “For example, **a trained machine-learning model can comprise a classifier** that is tasked with classifying a known input as one of multiple class labels, e.g., labeling the image of a cat or a dog.”

Ex. 1062, 217:16-218:10

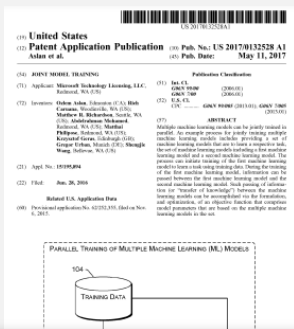


*Patent Owner's Expert
Eli Saber*

Q. Could it have been after 2017? It's possible that was the **first time your heard about teacher/student architecture**, it was after 2017?

A. **No. It would have been definitely before 2017.** It was...not, you know, something new.

Ex. 1062, 227:20-228:3



Aslan

“Because the smaller second model 102 take much less memory to maintain and can operate faster on less processing power at runtime, **the second model 102 can be a compressed form of the larger first model 100** such that the second model 102 can be more readily deployed on computing devices with limited resources (e.g., mobile devices, wearables, etc.).”

Ex. 1007, ¶44



Patent Owner's Expert
Eli Saber

Q. So is **Aslan** telling a POSITA in the pre 2017 time frame that its **joint training techniques can be used for model compression**?

A. **Yes.** He is making that statement, as well as stating **one example application is model compression.**

Ex. 1062, 239:16-22



Board

“Because the proposed combination involves **known elements performing their ordinary functions to produce a predictable result**, a POSITA would have appreciated that the proposed combination is proper.”

Institution Decision, 37

Motivation: Accuracy and/or Efficiency Gains



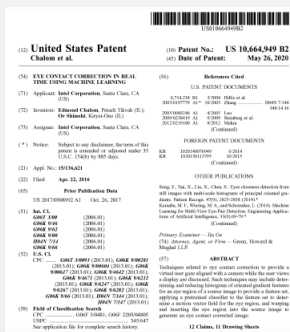
2

More Motivation: Accuracy and/or efficiency gains



“Nothing in Chalom indicates it is inaccurate or inefficient, and thus there is no reason to look to Aslan—and add another machine learning algorithm—to correct these falsified problems with Chalom.”

POR, 23



Chalom

“The techniques discussed herein may provide for high quality eye contact correction in real time.”

Ex. 1005, 9:50-52



**Patent Owner's Expert
Eli Saber**

A. So [Chalom] says... “The techniques described herein may provide for high quality eye contact correction in real time.” So **it’s essentially saying, well, it may provide for high quality eye contact correction, but there may be some error also now and then.** So it’s – it’s not saying it’s going to do it 100 percent of the time every single time...

Ex. 1061, 141:4-145:3

“system 1700 may be embodied in varying physical styles or form factors.”

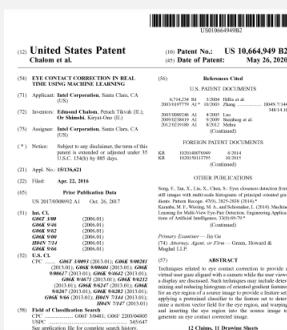
Ex. 1005, 22:32-33

“device 1800 may be implemented as a mobile computing device...”

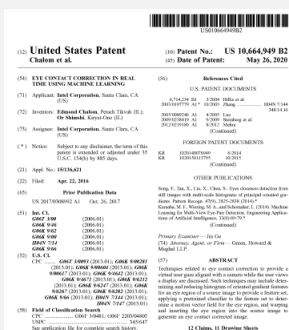
Ex. 1005, 22:40-42

“the interface may be any of a High-Definition Multimedia Interface, DisplayPort, wireless HDMI, and/or wireless HD compliant techniques.”

Ex. 1005, 20:3-6



Chalom



Chalom

“embodiments may include a training component (e.g., performed offline) and a real time component (e.g., performed during implementation or run time) that provides eye contact correction.”

Ex. 1005, 3:21-24

“The techniques discussed herein may provide for high quality eye contact correction in real time.”

Ex. 1005, 9:50-52

“histogram of oriented gradient features map 1301 may provide a robust appearance model for an eye region of an input image.”

Ex. 1005, 13:37-40



*Patent Owner's Expert
Eli Saber*

Q. Would **one of those problems in the field** be how to make these types of machine learning algorithms **more efficient**?

A. **Yeah, I would expect so, yes...** [I]f you have experience in machine learning, you would have experience in... designing them and **making more efficient or optimizing them in a sense.**

Ex. 1061, 180:14-21

Q. [W]hat kind of tradeoffs are involved with choosing sort of the number of parameters and selection of hyper parameters?

A. ...And usually the [] **time it takes to train, the size of the training database** that you have, **how long you're going to train**, how many epochs you're going to train, is the objective function convex or not...

Ex. 1061, 32:7-16



*Patent Owner's Expert
Eli Saber*

in order to continue to optimize the weights and parameters of the model. This process is then repeated multiple times using the entire training dataset to optimize the weights/parameters to yield higher accuracy results. Over time, human programmers can tweak the model, change/adjust parameters, and/or supply additional training data as needed to yield more accurate results. Once the training of the machine learning model is complete and all of the parameters and weights have been optimized for accuracy, it can then be utilized to perform the task that it was trained to do and can be further updated/upgraded in due time as additional training data becomes available.

Ex. 2008 (Saber decl.), ¶45

A. ... [T]he idea behind it, generally speaking, is you want to make changes to **improve accuracy or efficiency of a given model...**

Ex. 1061, 110:23-25



Petitioner's Expert
Lina Karam

Q. Do you agree or disagree that Chalom teaches a machine learning system that is fast enough to support 24 real-time implementation?

A. ... He mentions he wants to improve on things to be faster and to support real-time. **That said, one can always improve on real-time...**

Ex. 2009, 158:2-4

A. ... So it can support real-time implementation given the right resources. There's no limit on resources. But **if you want to limit to lower power and have limited resources for mobile, then you have to seek to improve on Chalom in order to have that...**

Ex. 2009, 159:7-12

A. ... You improve up to a certain degree. **Does this mean that you reached the perfect quality? Not necessarily...**

Ex. 1061, 110:23-25



“KSR did not impose a rigid requirement to identify both a problem to be solved in the art and a finite universe of potential options... We have similarly recognized that persons of ordinary skill have diverse motivations, including ‘[t]he normal desire ... to improve upon what is already generally known.’”

Jazz Pharmaceuticals Inc. v. Amneal Pharmaceuticals, LLC,
895 F.3d 1347, 1363 (Fed. Cir. 2018) (internal citations omitted)

REAL D

Patent Owner

“[A] POSITA would improve accuracy in other ways, none of which involve adding another algorithm. For example, a POSITA would improve an algorithm’s accuracy **by increasing the size/diversity of the training data or by changing certain features of the algorithm, such as its parameters or resolution.**”

POR, 23

“If a POSITA desired a more efficient way to deploy Chalom, they would reduce the size of the model **by having fewer trees and branches**, or compress the model by **selecting the most probable trees and branches...** **There is no reason to look for a second algorithm, or a teacher-student system,** to improve the efficiency of Chalom.”

POR, 26



“[O]ur case law does not require that a particular combination must be the preferred, or the most desirable, combination described in the prior art in order to provide motivation for the current invention. ‘[T]he question is whether there is something in the prior art as a whole to suggest the *desirability*, and thus the obviousness, of making the combination,’ **not whether there is something in the prior art as a whole to suggest that **the combination is the *most desirable* combination available.**”**

In re Fulton, 391 F.3d 1195, 1200 (Fed. Cir. 2004) (internal citations omitted)



*Patent Owner's Expert
Eli Saber*

Q. And [] in the pre-2017 timeframe, [] what is a teacher/student architecture?

A. Well... you have essentially [] **an architecture that provides perhaps [] an in-depth knowledge** in a sense or an in-depth training or an in-depth, you know, a set of algorithms. And **the other architecture might be a slightly smaller architecture or a slightly more efficient architecture** in a sense.

Ex. 1061, 182:2-24



*Petitioner's Expert
Lina Karam*

"[A]lthough **the teacher model can achieve a high accuracy**, it may not meet desired time and/or storage requirements for real-time deployment due to being too slow or too large... [T]he **student model, which is typically simpler or smaller and more suitable for deployment**, can be trained by **transferring the knowledge from the teacher ML model.**"

Ex. 1003, ¶52

“The described use of two machine learning algorithms allows an **improved balance** to be achieved **between accuracy of the adjustment process and speed of execution.**”

Ex. 1001, 2:17-19

“By arranging instead for the first reference data to be provided to a second machine learning algorithm... it is possible to **benefit to an extent from the high accuracy of the first machine learning algorithm whilst** also providing reference data that is less detailed and thus easier to **process efficiently** when performing the adjustment of the digital representation of the head region.”

Ex. 1001, 2:27-37



'985 Patent



“Petitioner’s proposed combination does not result in improved efficiency or accuracy. It does the exact opposite by **increasing complexity with more or larger models (which reduces efficiency)...**”

POR, 24-25



“[A] given course of action often has simultaneous advantages and disadvantages, and **this does not necessarily obviate motivation to combine.**”

Medichem, S.A. v. Rolabo, S.L., 437 F.3d 1157, 1165 (Fed. Cir. 2006)



“[O]ur case law does not require that a particular combination must be the preferred, or the most desirable, combination described in the prior art in order to provide motivation for the current invention. ‘[T]

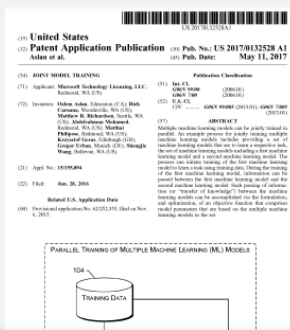
In re Fulton, 391 F.3d 1195, 1200 (Fed. Cir. 2004) (internal citations omitted)



*Petitioner's Expert
Lina Karam*

“And that’s why the teacher-student model would give you that advantage, because it takes the knowledge of the teacher now, learn, and passes it to the student. In addition, the student is taking the training data but also is taking that knowledge – the knowledge would bring the student to a level so that it can deal better with that training data. We have a different minimum. So if I can put you next to the minimum you need to be close to, you will converge faster to that. And that’s what the teacher does. It guides to the minimum that is -- gives the better accuracy.”

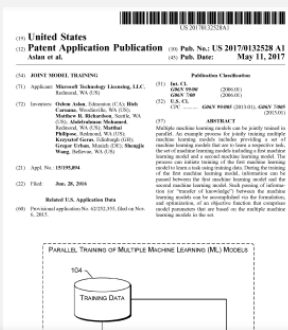
Ex. 2009, 178:14-25



Aslan

“By joint training the first and second models 100 and 102 using the techniques and systems described herein, **the second model 102 can be trained to mimic the much larger first model 100** (through learning how to approximate the function learned by the first model 100) **without significant loss in accuracy of the second model's 102 output.** Because the smaller second model 102 take much less memory to maintain and can operate faster on less processing power at runtime, **the second model 102 can be a compressed form of the larger first model 100** such that the second model 102 can be more readily deployed on computing devices with limited resources (e.g., mobile devices, wearables, etc.).”

Ex. 1007, ¶144



Aslan

“[T]his can allow for model compression, where the second type of machine learning model that **mimics the first type can**, at the completion of the joint training, **have a reduced size... without significant loss in accuracy** of the second model’s output.”

Ex. 1007, ¶4

“By joint training..., the second model 102 can be trained to **mimic the much larger first model 100... without significant loss in accuracy** of the second model’s 102 output.”

Ex. 1007, ¶44



*Petitioner's Expert
Lina Karam*

“That’s why this teacher-student combination is advantageous, because **you would have a bigger network learning and then passing the knowledge.** So this would avoid some of the problems related to the overfitting, underfitting, you know, that we discussed before.

A smaller network might not be able to learn from a complex dataset, because it doesn’t have the capabilities by itself. And that’s why **if you want to do low power and -- while maintaining the quality,** we need... this teacher- student combination is advantageous, because it helps us achieve that.”

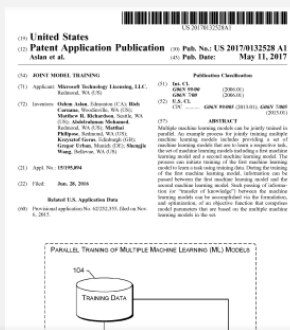
Ex. 2009, 175:24-176:15

“Machine learning models that are trained using the techniques described herein can **perform better (in terms of the accuracy of the model output)** than conventionally-trained machine learning models in some scenarios.”

Ex. 1007, ¶18

“the second (student) model 102 can comprise a **much smaller** machine learning model (e.g., a neural net with 1000 times fewer parameters than the first model 100) that **has the size and/or speed** that is advantageous at run-time in particular scenarios... the second model 102 can be trained to mimic the much larger first model 100 (through learning how to approximate the function learned by the first model 100) **without significant loss in accuracy of the second model's 102 output.**”

Ex. 1007, ¶44



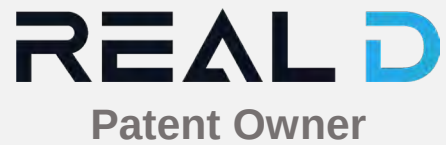
Aslan

Chalom-Aslan Combination is Sufficiently Explained



3

The Chalom-Aslan Combination is Sufficiently Explained



“Petitioner fails to sufficiently explain its Chalom-Aslan combination such that a POSITA could determine precisely what information passes between teacher and student models and how it operates.”

POR, 30



'985 Patent

“Any suitable machine learning technique may be applied, for example using a decision tree, a decision forest, a decision fern or an ensemble or combination thereof, or a neural network.”

Ex. 1001, 10:49-67

“By way of example, a suitable machine learning technique using a feature vector comprising features that are binary values derived by comparing a set of individual pixels or a linear combination thereof against a threshold, is disclosed in Ozuysal et al. “Fast Keypoint Recognition in Ten Lines of Code”, Computer Vision and Pattern Recognition, 2007, pp. 1-8, which is herein incorporated by reference in its entirety.”

Ex. 1001, 11:1-6



'985 Patent

“Reference to editing instructions herein is understood to encompass any data which can be used to define how a digital representation (e.g. image) of a head region should be adjusted to achieve a desired aim...”

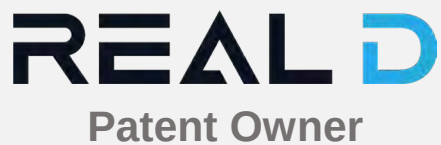
Ex. 1001, 18:21-24

“The method may use reference data generated using any of the methods of generating reference data disclosed herein or **may use reference data generated using other methods.”**

Ex. 1001, 11:1-6

“Any of various known techniques may be used to perform the updating of the second machine learning algorithm MLA2.”

Ex. 1001, 22:31-33



“[A] POSITA would not be motivated to combine Chalom and Aslan... because... Chalom performs a gaze correction task, while Aslan performs an image classification task.”

POR, 27



Patent Owner's Expert
Eli Saber

“**motion vector fields** represent two-dimensional displacement vectors... [and] are *typically* expressed via **decimal numbers**... However, **classes/labels** are **usually integers** (such as a number or index) and represent one dimension.”

Incorrect assumption

Ex. 2008, ¶119



*Patent Owner's Expert
Eli Saber*

Q. Are there other ways to express motion vectors besides decimal numbers?

A. ... **[Y]ou might express some with integers...**

Q. So... **you could express motion vectors as integers and not use interpolation.** Is that what you told me?

A. **Obviously, you could do that...**

Ex. 1062, 251:2-24



Petitioner's Expert
Lina Karam

“While Aslan describes the teacher-student architecture in the image-classification context, in my opinion, a PHOSITA would have been motivated to implement similar techniques for Chalom's image-editing context. This is at least because **determining, using machine learning, an optimal motion vector field for a given eye region is still an image-classification task.** Chalom's training process teaches all the possible motion vector fields to its MLAs, and **these possible motion vector fields**, although they will be used for image editing, **they are classes/labels themselves.** During runtime, **finding the optimal motion vector field corresponds to finding the best class/label and therefore involves a classification process.**”

Ex. 1003, ¶215



*Patent Owner's Expert
Eli Saber*

Q. So that Random Forest Classifier 306 that we see in Figure 3, Chalom Exhibit 1005, that is not a machine learning algorithm; is that correct?

A. ... [T]he **random forest classifier**... will do the **classification**... [I]t will **provide the appropriate motion vectors**.

Ex. 1061, 134:20-135:25

Q. Is a **random forest tree** a kind of algorithm?

A. Yeah..., it's a **classifier** type, obviously, algorithm that you would use... **mostly for classifications**.

Ex. 1061, 5:5-10



'985 Patent

3. The method of claim 1, wherein **the editing instructions comprise a displacement vector field** defining how the input patch is to be transformed.

Ex. 1001, Claim 3

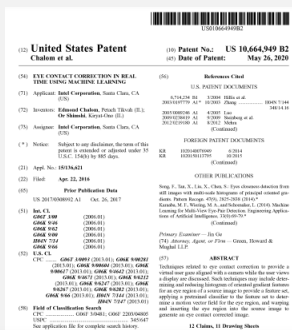
“By way of non-limitative example, the displacement vector field $\mathbf{F}$ may be derived as disclosed in Kononenko et al., ‘Learning To Look Up: Realtime Monocular Gaze Correction Using Machine Learning’, Computer Vision and Pattern Recognition, 2015, pp. 4667-4675, which is herein incorporated by reference in its entirety, wherein the displacement vector field $\mathbf{F}$ is referred to as a ‘flow field’.”

Ex. 1001, 9:6-18

REAL D
Patent Owner

“Petitioner’s combination also does not increase accuracy since it would result in horizontal gaze correction parameters being taught to vertical gaze correction models and vice versa. Chalom’s users may be looking below, above, right, or left of the camera, so **Chalom employs different models, each one to correct for a different potential gaze direction.**”

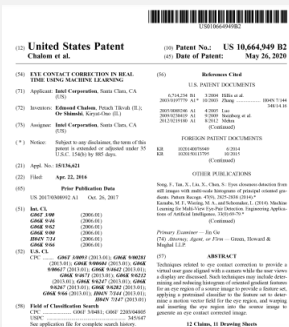
POR, 26



Chalom

“In other examples, a **single** eye contact correction module may implement (e.g., via loading from memory) different predetermined variables or data structures to provide **multiple eye contact corrections** each corresponding to **different relative positions** between camera 104 and display 102.”

Ex. 1005, 10:26-31



Chalom

“In other embodiments, system 200 may include multiple contact correction modules or contact correction module 204 may be capable of implementing eye contact correction for multiple orientations or relative positions between camera 104 and display 102.”

Ex. 1005, 10:10-14



**Patent Owner's Expert
Eli Saber**

Q. So turning back to Column 10, Line 10 to Line[] 14,... is that a disclosure of multiple machine learning algorithms in Chalom?

A. ... So **you could split them up** if you want that way, or **you could have them all in one bigger module**, so to speak, for multiple orientations.

Ex. 1061, 194:3-18

IV. Conclusion

PO Raises No Additional Rebuttal for Claims 2-14, 16-17, and 19

PO raises no additional rebuttal for:

- Ganin-based anticipation challenge to dependent claims 2-3, 5, 10, 12-14
- Ganin-Liu obviousness challenge to claims 16 and 19
- Chalom-Aslan obviousness challenge to claims 2, 3, 5, 6, 8-14, 16-17, and 19
- Chalom-Aslan-Grau obviousness challenge to claim 4
- Chalom-Aslan-Pace obviousness challenge to claim 7

Accordingly, claims 2-14, 16-17, and 19 rise and fall together with claims 1 and 18, and are therefore unpatentable for the reasons explained.

Reply to POR, 27-28

Conclusion: The Challenged Claims are Unpatentable

Ganin-led Grounds:

-- PO's arguments rely on imagined elements not required by the claims.

Chalom-Aslan-led Grounds:

-- Legally-sufficient express and other motivations existed. PO's other arguments misunderstand both the law and the facts.