

HONORIS CAUSA

**Acte d'investidura
del Dr. Antonio Torralba
com a doctor *honoris causa*
per la Universitat Politècnica de Catalunya • BarcelonaTech**



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Acte d'investidura
del Dr. Antonio Torralba
com a doctor *honoris causa*
per la Universitat Politècnica
de Catalunya · BarcelonaTech

11 de març de 2022



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Mitjana de persones a Cambridge (2003)

La imatge de la coberta està formada per la mitjana d'imatges que es fa a partir de la combinació de moltes imatges. Antonio Torralba va emprar aquesta tècnica per motivar l'estudi de models de context en visió artificial i per il·lustrar que la influència d'un objecte en una imatge s'estén més enllà dels límits de l'objecte. Abans de fer la mitjana, s'ha de traduir i escalar cada imatge de manera que l'objecte concret es trobi al centre de la imatge. Les imatges mitjanes alineades en un sol objecte que ocupa una petita part de la imatge poden revelar regions addicionals més enllà dels límits de l'objecte que proporcionen una estructura contextual significativa que li dona suport.

Més informació: <https://groups.csail.mit.edu/vision/torralbalab/#gallery>

Imprès en paper ecològic i sostenible procedent de boscos ben gestionats.
Servei de Comunicació de la UPC, 2022 (10243)

Índex / Índice / Table of contents

Ordre de l'acte d'investidura /	7
Orden del acto de investidura /	17
Order of the award ceremony /	27
Elogi dels mèrits del Dr. Antonio Torralba, Dr. Ferran Marqués /	9
Elogio de los méritos del Dr. Antonio Torralba, Dr. Ferran Marqués /	19
Oration for Prof Antonio Torralba by the sponsor Prof Ferran Marqués /	29
Discurs del nou doctor <i>honoris causa</i> , Dr. Antonio Torralba /	12
Discurso del nuevo doctor <i>honoris causa</i> , Dr. Antonio Torralba /	22
Acceptance speech by Prof Antonio Torralba /	32



Ordre de l'acte d'investidura

Benvinguda del rector de la Universitat Politècnica de Catalunya · BarcelonaTech, Dr. Daniel Crespo.

Lectura de l'Acord del Consell de Govern, a càrrec de la secretària general, Sra. Ana B. Cortinas.

Laudatio del padrí, Dr. Ferran Marqués.

Acte solemne d'investidura del Dr. Antonio Torralba com a doctor *honoris causa* per la Universitat Politècnica de Catalunya · BarcelonaTech.

Discurs del nou doctor *honoris causa*, Dr. Antonio Torralba.

Paraules del rector, Dr. Daniel Crespo.

Gaudeamus igitur, himne universitari. Harmonització de Joan Casulleras

Gaudeamus igitur iuvenes
dum sumus (bis),
post iucundam iuventutem,
post molestam senectutem
nos habebit humus (bis)

Ubi sunt qui ante nos
in mundo fuere (bis)
adeas ad inferos, transeas
ad superos
hos sivis videre (bis)

Vivat Academia, vivant
profesores, (bis)
vivat membrum quolibet,
vivant membra quaelibet
semper sint in flore (bis)



Elogi dels mèrits del Dr. Antonio Torralba

Dr. Ferran Marqués

Professor de la Universitat Politècnica de Catalunya · BarcelonaTech

El rector de la Universitat Politècnica de Catalunya, la meua universitat, a proposta de l'Escola de Telecomunicació de Barcelona, la meua escola, i del Centre de Formació Interdisciplinària Superior, institució amb la qual col·laboro amb orgull, em va encarregar que fes la *laudatio* del professor Antonio Torralba, persona que admiro des d'un gran nombre de punts de vista. Per tant, us podeu imaginar l'alegria i la responsabilitat que aquest encàrrec m'ha generat i que avui intentaré dur a terme.

El professor Antonio Torralba va rebre el títol d'Enginyer de Telecomunicació per l'Escola de Telecomunicació de Barcelona de la UPC l'any 1994 i el de doctor per l'Institut National Polytechnique de Grenoble l'any 2000. Des de l'any 2000 fins al 2007 va treballar al Massachusetts Institute of Technology (MIT), primer com a *postdoctoral associate* i després com a *research scientist*. L'any 2007 va ser contractat com a *assistant professor* i va obtenir la plaça d'*associate professor* el 2008. Des de l'any 2014 és catedràtic al Department of Electrical Engineering and Computer Science de l'MIT.

Durant els anys que fa que el professor Torralba treballa a l'MIT, ha obtingut diferents premis, com ara el National Science Foundation CAREER Award el 2008, el J. K. Aggarwal Prize de la International Association for Pattern Recognition el 2010, el Louis D. Smullin Award for Teaching Excellence el 2017 o l'Amazon Research Award el 2016.

L'any 2017 va ser nomenat *MIT Director* de l'MIT-IBM Watson AI Lab i l'any 2018, *Inaugural Director* del The MIT Quest for Intelligence. Aquesta és una iniciativa transversal a tot l'MIT que busca descobrir els fonaments de la intel·ligència i desenvolupar sota aquesta visió eines tecnològiques que puguin influir positivament en qualsevol aspecte de la societat.

Finalment, l'any 2020 va ser nomenat cap de la Faculty of Artificial Intelligence and Decision-Making, que és una nova unitat del Department of Electrical Engineering and Computer Science, al qual pertany. Aquesta unitat, basant-se en les arrels de l'enginyeria de telecomunicació i les ciències de la computació que té el Departament, pretén combinar i potenciar les àrees de l'aprenentatge automàtic, la intel·ligència artificial i la presa de decisions.

És evident que la llista de mèrits del professor Torralba que acabo d'esmentar ha estat incompleta a causa de la durada de l'acte i, per tant, segur que ha estat esbiaixada, ja que jo mateix, basant-me en unes quantes persones, n'he fet la tria. Aquest és un fet que no em preocupa, ja que, avui en dia, tota la informació quantitativa sobre el professor Torralba es pot trobar a internet. És per això que ara em voldria centrar en alguns aspectes addicionals de la figura del professor Torralba que crec que són importants per comprendre'n la magnitud i les raons per les quals avui rep aquest doctorat *honoris causa*.

Dins de la disciplina de la visió per ordinador, el treball del professor Antonio Torralba ha estat reconegut com a pioner i creador de les bases en l'àrea d'interpretació semàntica de l'escena. Les seves aportacions en l'estudi de l'escena des d'un punt de vista holístic han permès avançar en el desenvolupament de sistemes de reconeixement d'imatge que prenen en compte el context de l'escena, apropant les prestacions de les eines automàtiques a les dels éssers humans. En aquest àmbit, el seu treball ha esdevingut una referència fonamental i una lectura bàsica en la majoria dels cursos d'aquest tema arreu del món.

Ara bé, una de les característiques que fa especial la recerca del professor Torralba és el fet que ell treballa en la creació d'eines de visió per ordinador des d'una comprensió dels mecanismes de la percepció humana. D'aquesta manera, el professor Torralba estudia la psicofísica i realitza experiments que li permeten proposar i validar nous algoritmes i sistemes de visió per ordinador. La qualitat de la seva recerca en l'àmbit psicofísic l'ha dut a esdevenir també en aquesta àrea un referent mundial. El fet de poder combinar i complementar aquests dos punts de vista en la seva recerca ha fet que els treballs del professor Torralba siguin originals i innovadors i que la seva persona es visualitzi com un pont entre els dos àmbits.

A més, a cavall entre els aspectes de la visió per ordinador i la intel·ligència artificial, cal destacar l'interès del professor Torralba per caracteritzar de manera realista i exhaustiva els problemes que estudia. Això l'ha dut a construir i, més important, a fer públiques bases de dades anotades que han estat molt útils per als avenços de la comunitat. Ja des d'abans de l'any 2008, el professor Torralba ha promogut campanyes d'anotació de dades implicant-hi amplis recursos d'àmbit acadèmic i, com potser ens explicarà posteriorment el mateix professor Torralba, també d'àmbit familiar. I això ho ha fet mantenint sempre una visió crítica sobre quins són els límits de la validesa d'aquestes bases de dades i com han de generar-se, com demostra el seu treball

“Unbiased Look at Dataset Bias”, escrit en col·laboració amb el professor Alexei Efros.

És evident que tot el treball que he esmentat prèviament el van situar en una posició privilegiada quan es va produir l'eclosió dels sistemes d'aprenentatge profund (*deep learning*). Així, el professor Torralba ràpidament ha esdevingut també una figura molt rellevant en la branca de la intel·ligència artificial que treballa en els problemes de visió i en l'extrapolació de les eines que aquí es desenvolupen a altres àrees de la intel·ligència artificial.

D'aquesta manera, i com a evolució natural i integradora de totes les seves línies de recerca prèvies, el professor Torralba actualment centra la seva recerca en la fusió dels diferents aspectes de la percepció, en què extrapola les eines de la visió per ordinador i les contextualitza amb les altres modalitats perceptives, de manera que treballa en un entorn cada cop més multidisciplinari.

De fet, considero que, a més dels mèrits científics descrits prèviament, és tot aquest conjunt de valors que afloren en el seu treball els que han fet que l'Escola de Telecomunicació de Barcelona i el Centre de Formació Interdisciplinària Superior hagin proposat conjuntament el professor Torralba com a doctor *honoris causa* per la UPC. Crec que el seu enfocament obert vers la recerca, amb una cerca contínua de col·laboració generosa amb altres àmbits, i un intent constant d'impacte social són valors que les dues institucions senten com a propis, intenten inculcar als seus estudiants i agraeixen i reconeixen en el professor Torralba.

Però, a més, i ja per acabar aquesta lloança de la figura del professor Torralba, s'ha de remarcar que ha compaginat tot aquest impacte i visibilitat mundials amb un suport continu a la UPC, la seva *alma mater*. D'aquesta manera, el professor Torralba ha acollit ininterrompudament des de l'any 2011 estudiants de la UPC en el seu grup de recerca perquè poguessin dur a terme els seus treballs de fi d'estudis, tant de grau com de màster, alguns

dels quals posteriorment han passat a formar part del seu equip de recerca com a estudiants de doctorat. Aquest vincle sostingut amb la UPC, principalment mitjançant la seva relació amb l'Escola de Telecomunicació de Barcelona i el Centre de Formació

Interdisciplinària Superior, ha permès endegar nous projectes de col·laboració entre les dues institucions, com és ara el projecte Providence+, de monitoratge de les selves tropicals.

Discurs

del nou doctor *honoris causa*,

Dr. Antonio Torralba

Vull dedicar les meves primeres paraules a expressar el meu agraïment a la UPC per la concessió d'aquest títol de doctor *honoris causa*. La UPC, a banda de ser una universitat de gran prestigi, és també una institució molt especial per a mi. A la UPC em vaig graduar com a enginyer de telecomunicació i hi vaig estudiar per primer cop assignatures que es relacionaven amb el que a mi m'apassionava des de petit. També va ser a la UPC que vaig tenir l'oportunitat de fer recerca per primer cop, gràcies a una beca de col·laboració per a estudiants universitaris. Tinc també una relació especial amb la institució a través de diversos estudiants que han passat pel meu grup del MIT els darrers anys. Tot això fa que rebre aquesta distinció per part de la UPC tingui un gran significat per a mi.

És difícil dir què és el que em va motivar a ser investigador ni quan exactament va començar el meu interès per la intel·ligència artificial. Ja des de ben petit, m'atreia el món acadèmic i tenia clar que la meua vocació era la recerca. Quan tenia tretze anys, els meus pares em van regalar el meu primer ordinador, un Commodore 64, que encara conservo. Amb aquest ordinador vaig aprendre a programar i a dissenyar videojocs. Però ja aleshores el que em fascinava era la intel·ligència artificial. Al cap de poc temps, em van començar a interessar també totes les disciplines que tenien la més mínima relació amb el fet d'entendre en què

consisteix la intel·ligència. M'interessaven les matemàtiques, la física, la informàtica i l'electrònica. Però també m'apassionaven la filosofia, la neurociència i la psicologia. Feia servir els diners que guanyava fent classes de repàs per comprar-me llibres. Em fascinava llegir sobre investigadors distribuïts per tots els racons del món i somiava que un dia estaria entre ells. Però el món era molt gran, sobretot vist des de Mallorca en una època en què internet encara no existia. Als divuit anys, anar-me'n a Barcelona i començar la carrera a la UPC va suposar per mi l'inici d'una etapa d'exploració del món i d'introducció al món acadèmic.

Volia estudiar intel·ligència artificial, però aquells anys les carreres de ciència de dades i intel·ligència artificial no existien. Quan va arribar el moment d'escollir la carrera vaig decidir estudiar Enginyeria de Telecomunicació a la UPC i va ser després de començar la carrera quan se'm van obrir les portes al món de la recerca per primer cop. Concretament, el meu primer contacte amb la recerca va ser al laboratori de calorimetria del professor Vincenç Torra, al Departament de Física Aplicada, on vaig treballar en el desenvolupament de sistemes per experimentar amb materials amb memòria de forma, des de soldar components fins a programar sistemes de control de temperatura. L'últim any de carrera, a través de contactes proporcionats pel professor Joan Cabestany, me'n vaig anar a Grenoble, França, a fer el pro-

jecte de fi de carrera per dur a terme, finalment, recerca en visió per ordinador, que era el que realment volia fer. Em vaig quedar a Grenoble per fer el doctorat, sota la supervisió de Jeanny Héral, treballant en circuits neuromòrfics i visió. Després d'acabar la tesi, vaig decidir provar sort als Estats Units. L'any 2001 vaig aconseguir una plaça de postdoctorat al grup de William T. Freeman al MIT, on ara sóc professor.

La visió humana és sorprenent, capaç d'extreure informació del món que ens envolta fent servir només la llum que incideix en els objectes i que rebota cap als nostres ulls. La visió per ordinador estudia com es pot reproduir aquesta capacitat en els ordinadors i ha resultat que és una tasca molt complexa malgrat que els éssers humans, i la major part d'éssers vius, la realitzem sense cap esforç aparent.

Els darrers anys, la intel·ligència artificial, i la visió per ordinador en particular, han experimentat una revolució i formen part de molts dels sistemes que ens envolten. Quan fem una foto amb un telèfon, sembla obvi que serà capaç de detectar les cares presents per ajustar-hi automàticament la llum i enfocar la imatge. Com a conseqüència dels avenços dels darrers anys, avui dia entenem la visió com una enginyeria, l'objectiu de la qual és construir sistemes que puguin reconèixer objectes a partir d'imatges. Però la visió va ser a l'origen una ciència experimental. La visió és un fenomen natural del món biològic i l'objectiu de moltes generacions de científics ha consistit a esbrinar com funciona la intel·ligència i com la informació que rebem a través dels nostres sentits dona lloc a la nostra percepció conscient del món que ens envolta. Als orígens, l'estudi de la visió va ser multidisciplinària i va rebre contribucions de científiques i científics en física, matemàtiques, psicologia i neurociència. L'interès per la interdisciplinarietat ha estat una constant en la meua carrera.

Una de les àrees més importants dins de la visió per ordinador ha estat el reconeixement d'objectes. Els anys noranta, els models més populars de reconeixement d'objectes eren reduccionistes. Tractaven la imatge com un conjunt de regions que eren

processades de manera independent. L'exemple més clar són els sistemes de detecció de cares. Aquests sistemes funcionen trencant la imatge en milers de fragments i aplicant un classificador de cara a cada fragment de manera independent. L'èxit d'aquests sistemes va promoure l'aparició de centenars de nous algoritmes que feien servir aquest mateix mètode per detectar cotxes i vianants. Aquests sistemes no processaven l'escena com ho fa un ésser humà. No entenien el contingut de la imatge i no sabien fer res més que detectar un objecte concret basant-se en patrons visuals.

Durant la meua tesi, i en col·laboració amb la Dra. Aude Oliva, vam començar a treballar en un plantejament diferent en què l'escena en conjunt recuperava el protagonisme. El nostre objectiu era estudiar representacions que miraven la totalitat de la imatge i el resultat de les quals era extreure una descripció de l'escena més enllà dels objectes. El resultat d'aquest treball va ser el descriptor *gist*, publicat el 2001, que proporcionava una caracterització senzilla de la imatge i que estava inspirat en estudis que postulaven que el sistema visual humà extreia una descripció global de l'escena durant els primers 200 mil·lisegons. El *gist* de l'escena va aconseguir molta popularitat i va ser uns dels descriptors que més comunament es feren servir durant una dècada. Però la contribució més important del nostre treball, juntament amb altres investigadors, va ser inspirar una part de la comunitat a mirar el conjunt de la imatge. El descriptor *gist* i els models computacionals de context també els fem servir per modelitzar diferents aspectes de la visió humana, com la predicció de moviments oculars, i la memòria.

L'any 2003 vaig començar a treballar sobre models nous per introduir informació del context en el reconeixement d'objectes basats en mètodes d'aprenentatge. Encara que alguns models del context ja havien estat proposats anteriorment, hi havia pocs treballs i els que hi havia proposaven mètodes basats en regles, sense incloure-hi l'aprenentatge. Com a conseqüència, l'estudi del context havia quedat oblidat. Durant diversos anys vaig tenir la sort de col·laborar amb grans investigadors com Kevin Mur-

phy, Eric Sudderth, Rob Fergus, Ce Liu i Yair Weiss, amb els quals vaig publicar diversos articles sobre la detecció d'objectes en context, el *big data* i els models generadors d'escenes. Aquesta línia de treball va tenir molt impacte i va donar lloc a molts articles que van començar a introduir tota mena de models per incorporar informació del context a la detecció d'objectes.

En la dècada del 2000 al 2010, el reconeixement d'objectes (i no de cares) i la interpretació d'escenes van esdevenir protagonistes de la majoria de treballs en visió per ordinador. El 2007 vaig estudiar com es poden construir sistemes capaços de detectar centenars de tipus d'objectes diferents i compartir informació entre diferents categories per poder aprendre amb menys exemples. La comunitat de la visió per ordinador va començar a treballar en la interpretació d'escenes i en el reconeixement de centenars de categories d'objectes. A partir d'aquell moment va quedar clar que les bases de dades existents no eren suficients per entrenar els sistemes de visió que la comunitat científica començava a desenvolupar.

L' anotació d'imatges era una tasca àrdua en la comunitat investigadora. Calia un nou paradigma per crear bases de dades i anotar imatges. L'any 2006, en col·laboració amb Bryan Russell, vam crear LabelMe, una plataforma web oberta d'anotació d'imatges. L'objectiu era proporcionar una eina perquè els investigadors i investigadores poguessin anotar imatges sense haver d'escriure programari per fer-ho i també per poder sol·licitar ajuda d'altres anotadors en línia. LabelMe va ser un dels precursors del *crowdsourcing* per anotar imatges. Aquests treballs sobre la creació de bases de dades, juntament amb els treballs de context, van ser els que em van obrir les portes, el 2007, per obtenir la plaça de professor al MIT, on sóc encara.

Entre els anys 2005 i 2010 hi ha una explosió de bases de dades d'imatges quan la comunitat reconeix la importància de l'aprenentatge automàtic en la visió i la necessitat de tenir bases de dades prou grans per representar la riquesa del món visual. Altres avenços van venir al darrere i van aparèixer moltes bases

d'imatges, com ara Pascal i ImageNet, que van donar pas al *big data* en visió per ordinador. Les bases de dades també van revelar una sèrie de problemes, en particular l'aparició de biaixos a les imatges i a les anotacions. El 2011, en col·laboració amb Alyosha Efros, vam escriure un article crític sobre els conjunts d'imatges existents (incloent-hi els nostres) per motivar que la comunitat científica estudiés el problema.

El 2012 va arribar la revolució de l'aprenentatge profund (*deep learning*) i es van obrir moltes oportunitats per investigar aspectes de la visió que eren inabastables abans. El 2015 vam crear una base de dades de milions d'imatges anomenada Places, enfocada a l'estudi del reconeixement d'escenes. Places complementava la base de dades ImageNet, que tenia un enfocament centrat en la classificació d'objectes.

Les xarxes neuronals van demostrar que eren revolucionàries, però esdevenien opaques i no era fàcil descriure com resolien cadascuna de les tasques per a les quals eren entrenades. Tot i que les equacions que descriuen el funcionament de les xarxes neuronals són relativament senzilles, la naturalesa de la representació que calculen estava oculta. Una de les àrees en què el meu grup ha treballat durant els darrers anys ha estat la interpretació de les representacions apreses per les xarxes neuronals i el desenvolupament de tècniques per descriure com funcionen. Al meu grup hem treballat en el desenvolupament d'eines per analitzar com funcionen les xarxes neuronals i, en particular, per caracteritzar la representació interna que genera una xarxa neuronal quan analitza una imatge. Les tècniques que hem proposat estan inspirades en els mètodes que usen els neurocientífics per estudiar el funcionament del cervell humà. Aquests mètodes ens han permès estudiar molts tipus de xarxes neuronals i, en particular, les xarxes neuronals generatives.

Els darrers cinc anys, la meua recerca s'ha enfocat cap a la integració de la visió amb els altres sentits, com l'oïda i el tacte. El psicòleg James Gibson postulava que els sentits (la visió, l'oïda, el tacte i l'olfacte) ens proporcionen tanta informació sobre el

nostre entorn que no cal que el cervell faci computacions gaire complexes. Des de l'inici de la revolució de les xarxes neuronals, la comunitat científica ha centrat els seus esforços a construir xarxes neuronals cada cop més grans, amb bilions de connexions, que necessiten cada cop més recursos computacionals. Tot i això, l'estudi de sentits com el tacte encara són a les beceroles, malgrat que és possible que sigui essencial perquè els sistemes d'intel·ligència artificial puguin aprendre per si sols i connectar la seva experiència amb el món físic.

El camp de la intel·ligència artificial i la visió per ordinador estan vivint un moment de grans avenços, amb els quals s'obren les portes per a aplicacions revolucionàries en camps tan importants com el medi ambient i la salut, i perquè es desenvolupin tecno-

logies que impactaran en la societat, com ara els vehicles autònoms i la robòtica social, entre moltes altres. Alhora, els avenços en intel·ligència artificial també ens ajudaran a entendre la intel·ligència natural. Ha estat fascinant viure l'evolució de la intel·ligència artificial durant tots aquests anys com a investigador i haver compartit coneixements i experiències amb una extensa llista de mentors, col·laboradors i estudiants que han tingut un paper molt important en la meua trajectòria.

Per acabar, volia donar les gràcies a la meua família, als meus amics i amigues, a tots els que sou avui aquí per acompanyar-me en aquest moment. I una vegada més, vull agrair a la UPC que m'hagi atorgat aquest reconeixement.



Orden del acto de investidura

Bienvenida del rector de la Universitat Politècnica de Catalunya · BarcelonaTech, Dr. Daniel Crespo.

Lectura del Acuerdo del Consejo de Gobierno, a cargo de la secretaria general, Sra. Ana B. Cortinas.

Laudatio del padrino, Dr. Ferran Marqués.

Acto solemne de investidura del Dr. Antonio Torralba como doctor *honoris causa* por la Universitat Politècnica de Catalunya · BarcelonaTech.

Discurso del nuevo doctor *honoris causa*, Dr. Antonio Torralba.

Palabras del rector, Dr. Daniel Crespo.

Gaudeamus igitur, himno universitario. Armonización de Joan Casulleras

Gaudeamus igitur iuvenes
dum sumus (bis),
post iucundam iuventutem,
post molestam senectutem
nos habebit humus (bis)

Ubi sunt qui ante nos
in mundo fuere (bis)
adeas ad inferos, transeas
ad superos
hos sivos videre (bis)

Vivat Academia, vivant
profesores, (bis)
vivat membrum quolibet,
vivant membra quaelibet
semper sint in flore (bis)



Elogio de los méritos del Dr. Antonio Torralba

Dr. Ferran Marqués

Profesor de la Universitat Politècnica de Catalunya · BarcelonaTech

El rector de la Universitat Politècnica de Catalunya, mi universidad, a propuesta de la Escuela de Telecomunicación de Barcelona, mi escuela, y del Centro de Formación Interdisciplinaria Superior, institución con la que colaboro con orgullo, me encomendó que llevara a cabo la *laudatio* del profesor Antonio Torralba, persona que admiro desde muchos puntos de vista. Por tanto, pueden imaginar la alegría y la responsabilidad que este cometido me ha generado y que hoy intentaré llevar a buen término.

El profesor Antonio Torralba recibió el título de Ingeniero de Telecomunicación por la Escuela de Telecomunicación de Barcelona de la UPC en 1994 y el de doctor por el Institut National Polytechnique de Grenoble en el año 2000. Desde el año 2000 hasta el 2007 trabajó en el Massachusetts Institute of Technology (MIT), primero como *postdoctoral associate* y más tarde como *research scientist*. En el año 2007 fue contratado como *assistant professor* y obtuvo la plaza de *associate professor* en 2008. Desde el año 2014 es catedrático en el Department of Electrical Engineering and Computer Science del MIT.

Durante el tiempo que el profesor Torralba lleva trabajando en el MIT, ha obtenido varios premios, como el National Science Foundation CAREER Award en 2008, el J. K. Aggarwal Prize de la International Association for Pattern Recognition en 2010,

el Louis D. Smullin Award for Teaching Excellence en 2017 o el Amazon Research Award en 2016.

En 2017 fue nombrado *MIT Director* del MIT-IBM Watson AI Lab y en 2018, *Inaugural Director* del The MIT Quest for Intelligence, iniciativa transversal al conjunto de aquella institución, cuyo objeto es descubrir los fundamentos de la inteligencia y desarrollar bajo esa visión herramientas tecnológicas que puedan influir positivamente en cualquier aspecto de la sociedad.

Por último, en el año 2020 fue nombrado jefe de la Faculty of Artificial Intelligence and Decision-Making, que es una nueva unidad del Department of Electrical Engineering and Computer Science, del cual forma parte. Esta unidad, basándose en los fundamentos de la ingeniería de telecomunicación y las ciencias de la computación que posee el Departamento, pretende combinar y potenciar las áreas del aprendizaje automático, la inteligencia artificial y la toma de decisiones.

Resulta evidente que la lista de méritos del profesor Torralba que acabo de mencionar ha sido incompleta debido a la duración del acto y, por tanto, estoy convencido de que está sesgada, ya que yo mismo, con el apoyo de unas pocas personas, he realizado la elección. Este punto no me preocupa, puesto que, hoy en día, toda la información cuantitativa sobre el profesor Torralba pue-

de encontrarse en internet. Por todo ello quisiera centrarme ahora en algunos aspectos adicionales de la figura del profesor Torralba que creo son importantes para comprender su magnitud y las razones por las que hoy recibe este doctorado *honoris causa*.

Dentro de la disciplina de la visión por computador, el trabajo del profesor Antonio Torralba ha sido reconocido como pionero y creador de las bases en el área de interpretación semántica de la escena. Sus aportaciones al estudio de la escena desde un punto de vista holístico han permitido avanzar en el desarrollo de sistemas de reconocimiento de imagen que toman en cuenta el contexto de la escena, acercando las prestaciones de las herramientas automáticas a las de los seres humanos. En este ámbito, su trabajo se ha convertido en una referencia fundamental y una lectura básica en la mayoría de los cursos sobre este tema en todo el mundo.

Ahora bien, una de las características por las que la investigación del profesor Torralba puede considerarse especial es que él trabaja en la creación de herramientas de visión por computador desde una comprensión de los mecanismos de la percepción humana. De este modo, el profesor Torralba estudia la psicofísica y realiza experimentos que le permiten proponer y validar nuevos algoritmos y sistemas de visión por computador. Debido a la calidad de su investigación en el ámbito psicofísico se ha convertido también en esta área en un referente mundial. La combinación y complementación de estos dos puntos de vista en su investigación ha proporcionado a los trabajos del profesor Torralba originalidad e innovación y su persona se visualiza como un puente entre ambos ámbitos.

Asimismo, a caballo entre los aspectos de la visión por computador y la inteligencia artificial, cabe destacar el interés del profesor Torralba por la caracterización realista y exhaustiva de los problemas que estudia. Con ello ha construido y, más importante aún, ha hecho públicas bases de datos anotados de gran utilidad para los avances de la comunidad. Ya con anterioridad

al año 2008, el profesor Torralba fomentó campañas de anotación de datos implicando amplios recursos de origen académico y, como quizás nos explicará más tarde el mismo profesor Torralba, también familiar. Y todo ello manteniendo siempre una visión crítica sobre cuáles son los límites de la validez de estas bases de datos y cómo deben generarse, como demuestra su trabajo “Unbiased Look at Dataset Bias”, escrito en colaboración con el profesor Alexei Efros.

Es evidente que todo el trabajo que he mencionado previamente le situaron en una posición privilegiada cuando se produjo la eclosión de los sistemas de aprendizaje profundo (*deep learning*). Así, el profesor Torralba se ha convertido también en una figura de gran relevancia en el campo de la inteligencia artificial que trabaja en los problemas de visión y en la extrapolación de las herramientas que aquí se desarrollan a otras áreas de la inteligencia artificial.

De este modo, y como evolución natural e integradora de todas sus líneas de investigación previas, el profesor Torralba actualmente centra su investigación en la fusión de los diferentes aspectos de la percepción, en la que extrapola las herramientas de la visión por computador y las contextualiza con otras modalidades perceptuales, de tal forma que trabaja en un entorno cada vez más multidisciplinar.

Así pues, considero que, además de los méritos científicos descritos previamente, es todo este conjunto de valores que afloran en su trabajo los que han inducido a la Escuela de Telecomunicación de Barcelona y el Centro de Formación Interdisciplinaria Superior a proponer conjuntamente al profesor Torralba como doctor *honoris causa* por la UPC. Creo que su enfoque abierto hacia la investigación, con una búsqueda continua de colaboración generosa con otros ámbitos, y un intento constante de impacto social son valores que ambas instituciones sienten como propios, intentan inculcar a sus estudiantes y agradecen y reconocen en el profesor Torralba.

Pero, además, y ya para finalizar esta *laudatio* de la figura del profesor Torralba, cabe remarcar que ha compaginado todo este impacto y visibilidad mundiales con un apoyo continuo a la UPC, su *alma mater*. De este modo, el profesor Torralba ha acogido ininterrumpidamente desde el año 2011 a estudiantes de la UPC en su grupo de investigación para que pudieran llevar a cabo sus trabajos de fin de estudios, tanto de grado como de máster, pasando algunos de ellos a formar parte posteriormente

de su equipo de investigación como estudiantes de doctorado. Este vínculo sostenido con la UPC, principalmente mediante su relación con la Escuela de Telecomunicación de Barcelona y el Centro de Formación Interdisciplinaria Superior, ha permitido emprender nuevos proyectos de colaboración entre ambas instituciones, como por ejemplo el proyecto Providence+, de monitorización de las selvas tropicales.

Discurso del nuevo doctor *honoris causa*, Dr. Antonio Torralba

Quiero dedicar mis primeras palabras a expresar mi agradecimiento a la UPC por la concesión de este título de doctor *honoris causa*. La UPC, aparte de ser una universidad de gran prestigio, es también una institución muy especial para mí. En la UPC me gradué como ingeniero de telecomunicación y estudié por primera vez asignaturas que se relacionaban con lo que a mí me apasionaba desde pequeño. También fue en la UPC donde tuve la oportunidad de hacer investigación por primera vez, gracias a una beca de colaboración para estudiantes de carrera. Tengo también una relación especial con la institución a través de varios estudiantes que han pasado por mi grupo del MIT en los últimos años. Todo esto hace que recibir esta distinción por parte de la UPC tenga un gran significado para mí.

Es difícil decir qué es lo que me motivó a ser investigador ni cuándo exactamente empezó mi interés por la inteligencia artificial. Ya desde muy pequeño, me atraía el mundo académico y tenía claro que mi vocación era la investigación. Cuando tenía trece años, mis padres me regalaron mi primer ordenador, un Commodore 64, que aún conservo. Con ese ordenador aprendí a programar y a diseñar videojuegos. Pero ya entonces lo que me fascinaba era la inteligencia artificial. Al poco tiempo, me em-

pezaron a interesar también todas las disciplinas que tenían la más mínima relación con entender en qué consiste la inteligencia. Me interesaban las matemáticas, la física, la informática y la electrónica. Pero también me apasionaban la filosofía, la neurociencia y la psicología. Usaba el dinero que ganaba dando clases de repaso para comprarme libros. Me fascinaba leer sobre investigadores distribuidos por todos los rincones del mundo y soñaba con un día estar entre ellos. Pero el mundo era muy grande, sobre todo visto desde Mallorca en una época en la que internet aún no existía. A los dieciocho años, irme a Barcelona y empezar la carrera en la UPC supuso para mí el inicio de una etapa de exploración del mundo y de introducción al mundo académico.

Quería estudiar inteligencia artificial, pero en esos años las carreras de ciencia de datos e inteligencia artificial no existían. Cuando llegó el momento de elegir la carrera decidí estudiar Ingeniería de Telecomunicación en la UPC y fue después de empezar la carrera cuando se me abrieron las puertas al mundo de la investigación por primera vez. Concretamente, mi primer contacto con la investigación fue en el laboratorio de calorimetría del profesor Vincenç Torra, en el Departamento de Física Aplicada, donde trabajé en el desarrollo de sistemas para experimen-

tar con materiales con memoria de forma, desde soldar componentes hasta programar sistemas de control de temperatura. En el último año de carrera, a través de contactos proporcionados por el profesor Joan Cabestany, me fui a Grenoble, Francia, a hacer el proyecto final de carrera para llevar a cabo, por fin, investigación en visión por computador, que era lo que realmente quería hacer. Me quedé en Grenoble para hacer el doctorado, bajo la supervisión de Jeanny Hérault, trabajando en circuitos neuromórficos y visión. Tras acabar la tesis, decidí probar suerte en Estados Unidos. En el año 2001 conseguí una plaza de postdoctorado en el grupo de William T. Freeman en el MIT, donde ahora soy profesor.

La visión humana es sorprendente, capaz de extraer información del mundo que nos rodea usando solo la luz que incide en los objetos y que rebota en dirección a nuestros ojos. La visión por computador estudia cómo reproducir esta capacidad en computadores y ha resultado ser una tarea muy compleja pese a que los seres humanos, y la mayor parte de seres vivos, la realizamos sin esfuerzo aparente.

En los últimos años, la inteligencia artificial, y la visión por computador en particular, han experimentado una revolución y forman parte de muchos de los sistemas que nos rodean. Cuando hacemos una foto con un teléfono, nos parece obvio que será capaz de detectar las caras presentes para ajustar automáticamente la luz y enfocar la imagen. Como consecuencia de los avances de los últimos años, hoy en día entendemos la visión como una ingeniería, cuyo objetivo es construir sistemas que puedan reconocer objetos a partir de imágenes. Pero la visión fue en su origen una ciencia experimental. La visión es un fenómeno natural del mundo biológico y el objetivo de muchas generaciones de científicos ha consistido en averiguar cómo funciona la inteligencia y cómo la información que recibimos a través de nuestros sentidos da lugar a nuestra percepción consciente del mundo que nos rodea. En sus orígenes, el estudio de la visión fue multidisciplinar y recibió contribuciones

de científicas y científicos en física, matemáticas, psicología y neurociencia. El interés por la interdisciplinariedad ha sido una constante en mi carrera.

Una de las áreas más importantes dentro de la visión por computador ha sido el reconocimiento de objetos. En los años noventa, los modelos más populares de reconocimiento de objetos eran reduccionistas. Trataban la imagen como un conjunto de regiones que eran procesadas de forma independiente. El ejemplo más claro son los sistemas de detección de caras. Esos sistemas funcionan rompiendo la imagen en miles de trozos y aplicando un clasificador de cara en cada trozo de forma independiente. El éxito de estos sistemas promovió la aparición de cientos de nuevos algoritmos que usaban ese mismo método para detectar coches y peatones. Estos sistemas no procesaban la escena como lo hace un ser humano. No entendían el contenido de la imagen y no sabían hacer otra cosa más que detectar un objeto concreto basándose en patrones visuales.

Durante mi tesis, y en colaboración con la Dra. Aude Oliva, comenzamos a trabajar en un planteamiento distinto en el que la escena en su conjunto recuperaba el protagonismo. Nuestro objetivo era estudiar representaciones que miraban la imagen en su totalidad y cuyo resultado era extraer una descripción de la escena más allá de los objetos. El resultado de este trabajo fue el descriptor *gist*, publicado en 2001, que proporcionaba una caracterización sencilla de la imagen y que estaba inspirado en estudios que postulaban que el sistema visual humano extraía una descripción global de la escena durante los primeros 200 milisegundos. El *gist* de la escena alcanzó mucha popularidad y fue uno de los descriptores más comúnmente usados durante una década. Pero la contribución más importante de nuestro trabajo, junto al de otros investigadores, fue inspirar a una parte de la comunidad a mirar la imagen en su conjunto. El descriptor *gist* y los modelos computacionales de contexto también los usamos para modelizar diferentes aspectos de la visión humana, como la predicción de movimientos oculares, y la memoria.

En el año 2003, comencé a trabajar sobre nuevos modelos para introducir información del contexto en el reconocimiento de objetos basados en métodos de aprendizaje. Aunque algunos modelos del contexto ya habían sido propuestos anteriormente, existían pocos trabajos y los que había proponían métodos basados en reglas, sin incluir aprendizaje. Como consecuencia, el estudio del contexto había caído en el olvido. Durante varios años tuve la suerte de colaborar con grandes investigadores como Kevin Murphy, Eric Sudderth, Rob Fergus, Ce Liu y Yair Weiss, con quienes publiqué varios artículos sobre la detección de objetos en contexto, el *big data* y los modelos generativos de escenas. Esta línea de trabajo tuvo mucho impacto y dio lugar a muchos artículos que empezaron a introducir todo tipo de modelos para incorporar información del contexto en la detección de objetos.

En la década del 2000 al 2010, el reconocimiento de objetos (y no de caras) y la interpretación de escenas se volvieron protagonistas de la mayoría de trabajos en visión por computador. En el 2007 estudié cómo construir sistemas capaces de detectar cientos de tipos de objetos distintos y de compartir información entre diferentes categorías para poder aprender con menos ejemplos. La comunidad de la visión por computador comenzó a trabajar en la interpretación de escenas y en el reconocimiento de cientos de categorías de objetos. A partir de ese momento quedó claro que las bases de datos existentes no eran suficientes para entrenar los sistemas de visión que la comunidad científica estaba empezando a desarrollar.

La anotación de imágenes era una tarea ardua en la comunidad investigadora. Hacía falta un nuevo paradigma para la creación de bases de datos y la anotación de imágenes. En el año 2006, en colaboración con Bryan Russell, creamos LabelMe, una plataforma web abierta de anotación de imágenes. El objetivo era proporcionar una herramienta para que los investigadores e investigadoras pudiesen anotar imágenes sin tener que escribir software para hacerlo y también para poder solicitar ayuda de otros anotadores *online*. LabelMe fue uno de los precursores del

crowdsourcing para la anotación de imágenes. Estos trabajos sobre la creación de bases de datos, junto con los trabajos de contexto, fueron los que me abrieron las puertas, en el 2007, para obtener la plaza de profesor en el MIT, donde aún sigo.

Entre los años 2005 y 2010 hay una explosión de bases de datos de imágenes cuando la comunidad reconoce la importancia del aprendizaje automático en la visión y la necesidad de tener bases de datos su,cientemente grandes como para representar la riqueza del mundo visual. Otros avances siguieron a este y aparecieron muchas bases de imágenes, como Pascal e ImageNet, que dieron paso al *big data* en visión por computador. Las bases de datos también revelaron una serie de problemas, en particular la aparición de sesgos en las imágenes y en las anotaciones. En el 2011, en colaboración con Alyosha Efros, escribimos un artículo crítico sobre los conjuntos de imágenes existentes (incluyendo los nuestros) para motivar el estudio del problema por parte de la comunidad científica.

En el 2012 llegó la revolución del aprendizaje profundo (*deep learning*) y se abrieron muchas oportunidades para investigar aspectos de la visión que resultaban inalcanzables antes. En 2015 creamos una base de datos de millones de imágenes llamada Places, enfocada al estudio del reconocimiento de escenas. Places complementaba la base de datos ImageNet, que tenía un enfoque centrado en la clasificación de objetos.

Las redes neuronales demostraron ser revolucionarias, pero resultaban ser opacas y no era fácil describir cómo resolvían cada una de las tareas para las cuales eran entrenadas. Aunque las ecuaciones que describen el funcionamiento de las redes neuronales son relativamente sencillas, la naturaleza de la representación que calculan permanecía oculta. Una de las áreas en las que mi grupo ha trabajado durante los últimos años ha sido la interpretación de las representaciones aprendidas por las redes neuronales y el desarrollo de técnicas para describir cómo funcionan. En mi grupo hemos trabajado en el desarrollo de herra-

mientas para analizar cómo funcionan las redes neuronales y, en particular, para caracterizar la representación interna que genera una red neuronal cuando analiza una imagen. Las técnicas que hemos propuesto están inspiradas de los métodos que usan los neurocientíficos para estudiar el funcionamiento del cerebro humano. Estos métodos nos han permitido estudiar muchos tipos de redes neuronales y, en particular, las redes neuronales generativas.

Los últimos cinco años, mi investigación se ha enfocado hacia la integración de la visión con los otros sentidos, como el oído y el tacto. El psicólogo James Gibson postulaba que los sentidos (la visión, el oído, el tacto y el olfato) nos proporcionan tanta información sobre nuestro entorno que no es necesario que el cerebro realice computaciones muy complejas. Desde el inicio de la revolución de las redes neuronales, la comunidad científica ha centrado sus esfuerzos en construir redes neuronales cada vez más grandes, con billones de conexiones, que necesitan cada vez más recursos computacionales. Sin embargo, el estudio de sentidos como el tacto están aún en su infancia, pese a que es posible que sea esencial para que los sistemas de inteligencia

artificial puedan aprender por sí solos y conectar su experiencia con el mundo físico.

El campo de la inteligencia artificial y la visión por computador están viviendo un momento de grandes avances, con los que se abren las puertas para aplicaciones revolucionarias en campos tan importantes como el medio ambiente y la salud, y para que se desarrollen tecnologías que impactarán la sociedad, como los vehículos autónomos y la robótica social, entre muchas otras. A su vez, los avances en inteligencia artificial también nos ayudarán a entender la inteligencia natural. Ha sido fascinante vivir la evolución de la inteligencia artificial durante todos estos años como investigador y haber compartido conocimientos y experiencias con una extensa lista de mentores, colaboradores y estudiantes que han tenido un papel muy importante en mi trayectoria.

Para acabar, quería dar las gracias a mi familia, a mis amigos y amigas, a todos los que estáis hoy aquí por acompañarme en este momento. Y una vez más, quiero agradecer a la UPC que me haya otorgado este reconocimiento.



Order of the award ceremony

Welcome from the rector of the Universitat Politècnica de Catalunya · BarcelonaTech,
Prof Daniel Crespo.

Reading of the Governing Council's decision by the general secretary, Ms Ana B. Cortina.

Oration for Prof Antonio Torralba by the sponsor, Prof Ferran Marqués.

Conferral of the honorary doctorate in Prof Antonio Torralba
by the Universitat Politècnica de Catalunya · BarcelonaTech.

Acceptance speech by Prof Antonio Torralba.

Speech by the rector, Prof Daniel Crespo.

Gaudeamus igitur, university hymn. Arranged by Joan Casulleras

Gaudeamus igitur iuvenes
dum sumus (bis),
post iucundam iuventutem,
post molestam senectutem
nos habebit humus (bis)

Ubi sunt qui ante nos
in mundo fuere (bis)
adeas ad inferos, transeas
ad superos
hos sivos videre (bis)

Vivat Academia, vivant
profesores, (bis)
vivat membrum quolibet,
vivant membra quaelibet
semper sint in flore (bis)



Oration for Prof Antonio Torralba

*Prof Ferran Marqués
Professor at the Universitat Politècnica de Catalunya · BarcelonaTech*

The rector of the Universitat Politècnica de Catalunya, my university, at the proposal of the Barcelona School of Telecommunications Engineering, my school, and the Interdisciplinary Higher Education Centre, an institution with which I am proud to collaborate, asked me to give the oration for Professor Antonio Torralba, a person I admire in many ways. So you can imagine my pleasure and the sense of responsibility I felt to be entrusted with the task that I am undertaking today.

Professor Antonio Torralba received a degree in Telecommunications Engineering from the Barcelona School of Telecommunications of the UPC in 1994 and a doctoral degree from the Grenoble Institute of Technology in 2000. From 2000 to 2007, he worked at the Massachusetts Institute of Technology (MIT), first as a postdoctoral associate and then as a research scientist. In 2007, he was hired as an assistant professor and he obtained the post of associate professor in 2008. Since 2014, he has been a full professor in the Department of Electrical Engineering and Computer Science at MIT.

Over the years that Professor Torralba has worked at MIT, he has received various awards, such as the National Science Foundation CAREER Award in 2008, the J. K. Aggarwal Prize from the International Association for Pattern Recognition in 2010,

the Louis D. Smullin Award for Teaching Excellence in 2017 and the Amazon Research Award in 2016.

In 2017, he was appointed MIT director of the MIT–IBM Watson AI Lab and, in 2018, inaugural director of the MIT Quest for Intelligence. This is a cross-disciplinary initiative across MIT that seeks to discover the principles of intelligence and develop under this vision technological tools that could have a positive influence on any aspect of society.

Finally, in 2020, he was appointed head of the Faculty of Artificial Intelligence and Decision–Making, which is a new unit of the Department of Electrical Engineering and Computer Science to which he belongs. This unit, which is based on the Department’s roots in telecommunications engineering and computer sciences, is designed to combine and boost the areas of automatic learning, artificial intelligence and decision-making.

Clearly, the list of Professor Torralba’s merits that I have just given is incomplete because of the time available at this event. It is probably biased as I chose the items, using information from several people. However, I don’t think this is a problem, as all quantitative information on Professor Torralba can be found on the internet. For this reason, now I would like to focus on some

additional aspects of the figure of Professor Torralba that I believe are important to understand the magnitude of his work and the reasons why he is receiving this honorary doctoral degree today.

Within the discipline of computer vision, Professor Antonio Torralba has been recognised for instigating and establishing the foundations of the area of semantic scene interpretation. His contributions to the study of scenes from a holistic perspective have led to advances in the development of image recognition systems that consider the context of a scene and bring the features of machine tools closer to those of humans. In this area, his work has become an essential reference and basic reading in most courses on this topic worldwide.

However, one of the characteristics that makes Professor Torralba's research special is the fact that he works to create computer vision tools from an understanding of the mechanisms of human perception. In this way, Professor Torralba studies psychophysics and carries out experiments that enable him to propose and validate new algorithms and computer vision systems. The quality of his research in the psychophysical area has made him a world leader in this area too. The fact that he can combine and complement these two perspectives in his research has meant that Professor Torralba's studies are original and innovative, and he is seen as a bridge between both areas.

In addition, halfway between aspects of computer vision and artificial intelligence, Professor Torralba is interested in characterising the problems that he studies in a realistic, exhaustive way. This has led him to construct and, more importantly, to publish annotated databases that have been very useful for the advances of the community. Since before 2008, Professor Torralba has promoted data annotation campaigns using broad research in the academic area and, as he will perhaps explain later, in the family sphere. At the same time, he has always maintained a

critical view of the limits of the validity of these databases and how they should be generated, as shown in his work "Unbiased Look at Dataset Bias", written in collaboration with Professor Alexei Efros.

Clearly, all of the work that I have mentioned put him in a privileged position when deep learning systems emerged. Professor Torralba rapidly became a highly relevant figure in the branch of artificial intelligence that focuses on vision problems and on the extrapolation of tools that are developed in this field to other areas of artificial intelligence.

In this way, and as a natural, integrating evolution of all his previous lines of research, Professor Torralba currently focuses his work on the fusion of various aspects of perception. He extrapolates computer vision tools and contextualises them with other perceptive modalities, so he works in an increasingly multidisciplinary environment.

In fact, I consider that, in addition to the scientific merits that I have described, it is the entire set of values that emerge from his work that led to the Barcelona School of Telecommunications Engineering and the Interdisciplinary Higher Education Centre jointly proposing Professor Torralba for an honorary doctoral degree from the UPC. I believe that his open approach to research, continuous search for generous collaboration with other areas and constant focus on social impact are values that the two institutions recognise as their own, try to inculcate in their students and appreciate and recognise in Professor Torralba.

However, in addition and to conclude this laudation of the figure of Professor Torralba, I should mention that he has combined this impact and global visibility with continuous support for the UPC, his alma mater. Since 2011, Professor Torralba has uninterruptedly welcomed UPC students into his research group so that they can carry out their bachelor's and master's theses.

Some of these students have gone on to form part of his research team as doctoral students. This link maintained with the UPC, principally through his relationship with the Barcelona School of Telecommunications Engineering and the Interdisciplinary

Higher Education Centre, has enabled new collaboration projects between the two institutions, such as the Providence+ project on monitoring tropical forests.

Acceptance speech by Prof Antonio Torralba

First of all, I would like to thank the UPC for awarding me this honorary doctoral degree. In addition to being a university of great prestige, the UPC is a very special institution for me. At the UPC, I graduated as a telecommunications engineer and for the first time I studied subjects related to what I had been passionate since I was young. It was also at the UPC that I had the opportunity to do research for the first time, thanks to a collaboration grant for degree students. I also have a special relationship with the institution through the various students who have passed through my MIT group in recent years. All of this means that receiving this recognition from the UPC is of great significance to me.

It is hard to say what motivated me to be a researcher or exactly when my interest in artificial intelligence began. From a very young age, I was attracted to the academic world and I was sure that I wanted to work in research. When I was 13 years old, my parents gave me my first computer, a Commodore 64 that I still have. I learnt to program and design video games with that computer. But even then artificial intelligence was what fascinated me most. Shortly afterwards, I also became interested in all the disciplines that had even the slightest relation to understanding what intelligence consists of. I was interested in mathematics, physics, computer science and electronics. But I was also passionate about philosophy, neuroscience and psychology. I used the money I earned teaching remedial classes to buy books. I was fas-

inated to read about researchers all over the world and I dreamt that one day I would be one of them. But the world was very big, particularly when seen from Mallorca in a period before the internet existed. At the age of 18, going to Barcelona and starting a degree at the UPC was for me the start of a period for exploring the world and being introduced to academia.

I wanted to study artificial intelligence but degrees in data science and artificial intelligence did not exist then. When the time came to choose a degree, I decided to study telecommunications engineering at the UPC. The doors to the world of research first opened to me after I had started my degree. My first contact with research was in the calorimetry laboratory of Professor Vincenç Torra, in the Department of Applied Physics, where I worked on the development of systems to experiment with shape memory materials, from soldering components to programming temperature control systems. In the last year of the degree, through contacts provided by Professor Joan Cabestany, I went to Grenoble, France, to prepare my final thesis and, finally, to research computer vision, which is what I really wanted to do. I stayed in Grenoble to do my doctoral degree under the supervision of Jeanny Hérault, working on neuromorphic circuits and vision. Having completed my thesis, I decided to try my luck in the United States. In 2001, I gained a postdoctoral position in William T. Freeman's group at MIT, where I am now a professor.

Human vision is surprising. It is capable of extracting information about the world that surrounds us using only the light that falls on objects and is reflected in the direction of our eyes. Computer vision studies how to reproduce this capacity in computers. It has ended up being a very complex task, despite the fact that humans and most living beings see without apparent effort.

In recent years, artificial intelligence and computer vision, in particular, have undergone a revolution and now form part of many of the systems that surround us. When we take a photograph with a phone, it seems evident to us that the phone will be able to detect faces to automatically adjust the light and focus the image. As a result of advances in recent years, today we understand vision as engineering, the aim of which is to construct systems that can recognise objects from images. However, vision was originally an experimental science. Vision is a natural phenomenon in the biological world, and the objective of many generations of scientists has been to determine how intelligence functions and how the information that we receive through our senses leads to our conscious perception of the world around us. In its origins, the study of vision was multidisciplinary and it received contributions from scientists in physics, mathematics, psychology and neuroscience. The interest in interdisciplinarity has been a constant factor in my career.

One of the most important areas within computer vision has been object recognition. In the 1990s, the most popular models of object recognition were reductionist. They approached the image as a set of regions that were processed independently. The clearest example are face detection systems. These systems function by breaking the image into thousands of pieces and applying a face classifier to each piece independently. The success of these systems promoted the emergence of hundreds of new algorithms that used this method to detect cars and pedestrians. These systems did not process the scene as a human does. They did not understand the content of the image and they did not

know how to do anything other than detect a specific object based on visual patterns.

During my thesis, in collaboration with Dr Aude Oliva, we began to work on a different approach in which the scene as a whole recovered its leading role. Our aim was to study representations that looked at the image in its totality, to extract a description of the scene beyond the objects. The result of this work was the gist descriptor, published in 2001, which provided a simple characterisation of the image and was inspired by studies that proposed that human vision extracted a global description of the scene during the first 200 milliseconds. The gist of the scene became very popular and was one of the most commonly used descriptors for a decade. However, the most important contribution of our work, along with that of other researchers, was to inspire part of the community to look at the image as a whole. We also used the gist descriptor and computational models of the context to model various aspects of human vision, such as predicting eye movements, and memory.

In 2003, I began to work on new models to introduce information from the context into object recognition, based on learning methods. Although some models of the context had been proposed in the past, few studies existed and those that did had proposed rule-based methods and did not include learning. As a result, the study of the context had fallen into oblivion. For several years, I was lucky to collaborate with great researchers such as Kevin Murphy, Eric Sudderth, Rob Fergus, Ce Liu and Yair Weiss, with whom I published several articles on object detection in context, big data and scene generative models. This area of work had a considerable impact and led to many articles that started to introduce all kinds of models to incorporate information on the context into object detection.

In the decade from 2000 to 2010, object (not facial) recognition and scene interpretation again became the main focus of most

works in computer vision. In 2007, I studied how to construct systems that could detect hundreds of different types of objects and share information between categories, to be able to learn with fewer examples. The computer vision community began to work on scene interpretation and on the recognition of hundreds of categories of objects. From this point, it was clear that the existing databases were insufficient to train the vision systems that the scientific community was starting to develop.

Image annotation was an arduous task in the research community. A new paradigm was required for the creation of databases and image annotation. In 2006, in collaboration with Bryan Russell, we created LabelMe, an open web platform for image annotation. The aim was to provide a tool so that researchers could annotate images without having to write software to do this, and to be able to request help from other annotators online. LabelMe was one of the precursors of crowdsourcing for image annotation. These studies on the creation of databases, together with studies on context, were what opened the doors for me to obtain, in 2007, a post as professor at MIT, where I still work today.

Between 2005 and 2010, there was an explosion of image databases when the community recognised the importance of machine learning in vision and the need to have sufficiently large databases to represent the wealth of the visual world. Other advances followed this and many image databases emerged, such as Pascal and ImageNet, which led to big data in computer vision. The databases also revealed a series of problems, particularly the emergence of biases in images and annotations. In 2011, in collaboration with Alyosha Efros, we authored a critical article about existing sets of images (including ours) to stimulate the study of the problem by the scientific community.

In 2012, the deep learning revolution arrived and many opportunities emerged to research aspects of vision that had previously been unattainable. In 2015, we created a database of millions of

images called Places that was focused on the study of scene recognition. Places complemented the ImageNet database, whose approach was focused on object classification.

Neural networks were revolutionary, but they were opaque and it was not easy to describe how they resolved each task for which they were trained. Although the equations that describe the functioning of neural networks are relatively simple, the nature of the representation that they calculate remained hidden. One of the areas in which my group has worked in recent years has been the interpretation of representations learnt by neural networks and the development of techniques to describe how they function. In my group, we have worked on the development of tools to analyse how neural networks function and specifically to characterise the internal representation that a neural network generates when it analyses an image. The techniques that we have proposed are inspired by the methods that neuroscientists use to study how the human brain functions. These methods have enabled us to study many types of neural networks and particularly generative neural networks.

In the last five years, my research has focused on the integration of vision with other senses such as hearing and touch. Psychologist James Gibson proposed that the senses (vision, hearing, touch and smell) provide so much information on our environment that the brain does not need to perform very complex calculations. Since the start of the neural network revolution, the scientific community has focused its efforts on constructing ever larger neural networks, with billions of connections that require increasing computational resources. However, the study of senses such as touch is still in its infancy, even though it may be essential for artificial intelligence systems to learn for themselves and connect their experience to the physical world.

The field of artificial intelligence and computer vision is experiencing a period of great advances, in which doors are being

opened for revolutionary applications in fields as important as the environment and health and for which technologies are being developed that will have an impact on society, such as autonomous vehicles and social robotics, among many others. In turn, advances in artificial intelligence help us to understand natural intelligence. It has been fascinating to experience the evolution of artificial intelligence during all these years as a researcher and to have shared knowledge and experience with a

long list of mentors, collaborators and students who have played a very important role in my career.

To conclude, I would like to thank my family, my friends and everyone who is here today to accompany me at this event. And once more, I would like to thank the UPC for giving me this recognition.











**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**