

Transfer Learning

conceptos, casos de estudio, tutorial

Miguel Ángel Martínez del Amor

Dpto. Ciencias de la Computación e Inteligencia Artificial

Universidad de Sevilla

Agenda

- Conceptos básicos
- Casos de estudio:
 - Tag extraction
 - Glaucoma
- Tutorial con Keras:
 - Sesión 1
 - Sesión 2
 - Ejercicios

Agenda

- **Conceptos básicos**
- Casos de estudio:
 - Tag extraction
 - Glaucoma
- Tutorial con Keras:
 - Sesión 1
 - Sesión 2
 - Ejercicios

Transfer Learning

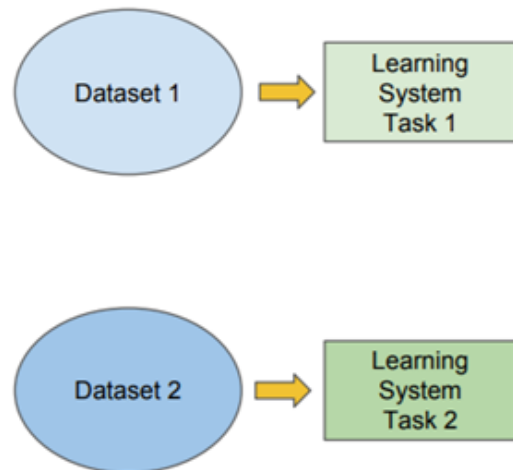
- Supongamos **nuestra tarea (T1)** es identificar objetos en imágenes de un restaurante (**dominio 1**). Entrenamos un modelo para ello.
- Supongamos **nueva tarea (T2)** es identificar objetos en imágenes de un parque o un café (**dominio 2**). Si aplicamos el modelo entrenado para T1, veremos una degradación de rendimiento para el dominio 2.
- **Transferencia de aprendizaje** debería permitirnos reutilizar el conocimiento de tareas previamente aprendidas y aplicarlas a nuevas.



ML tradicional vs Transfer Learning

Traditional ML

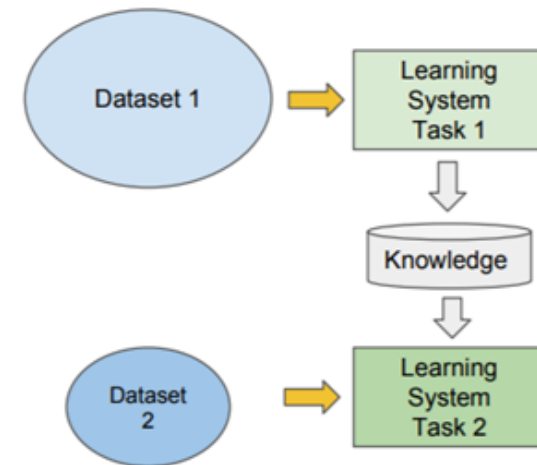
- Isolated, single task learning:
 - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks



vs

Transfer Learning

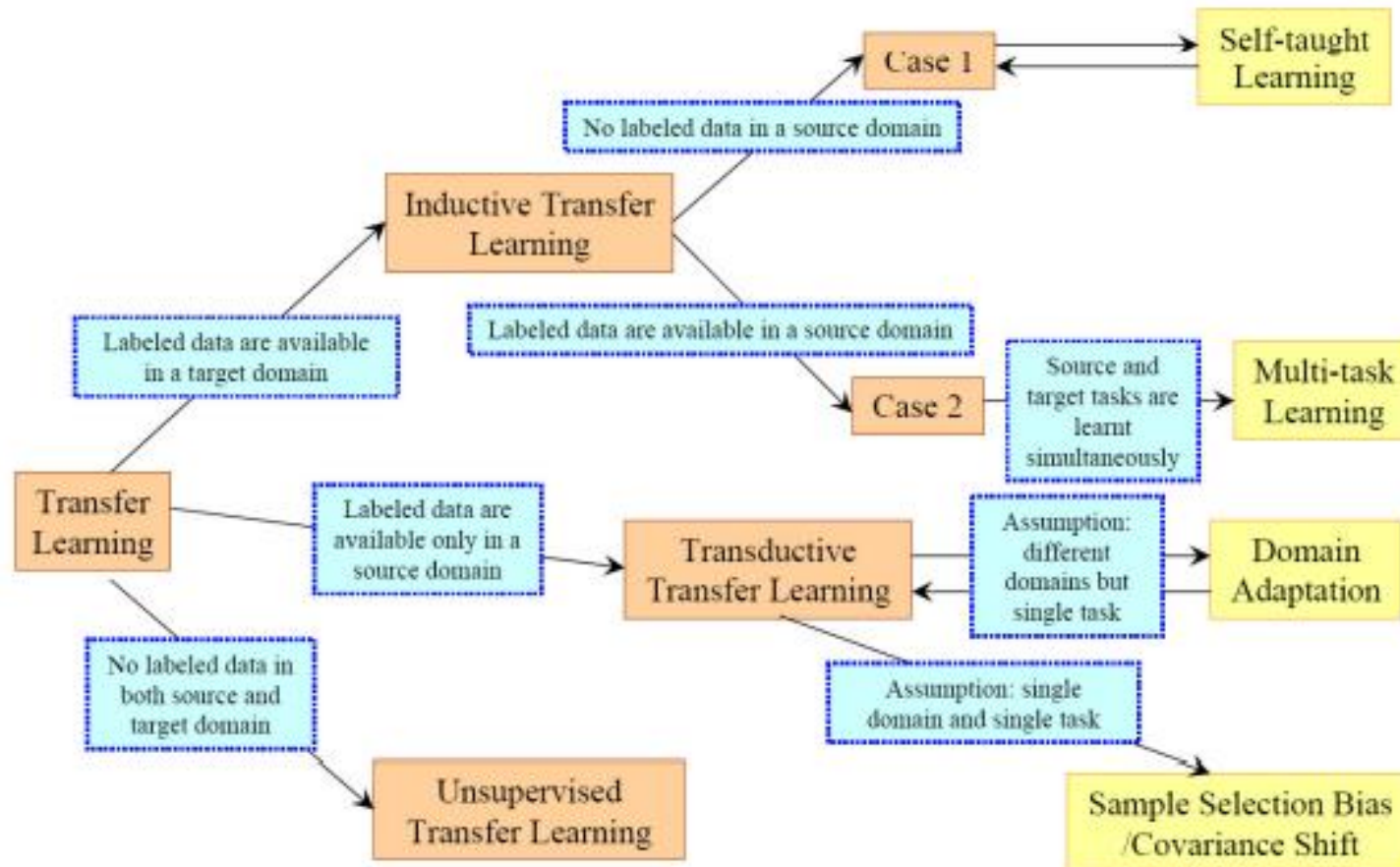
- Learning of a new tasks relies on the previous learned tasks:
 - Learning process can be faster, more accurate and/or need less training data



Transfer Learning

- **Transfer Learning:** habilidad de reutilizar conocimiento existente sobre una tarea origen en otras tareas objetivo.
 - ¿Qué transferir?
 - Identificar qué porción de conocimiento es específico de nuestro origen, y qué tiene en común el origen con el objetivo.
 - ¿Cuándo transferir?
 - Hay algunos escenarios donde transferir conocimiento es contraproducente.
 - ¿Cómo transferir?
 - Buscar los cambios en los algoritmos y las diferentes técnicas existentes para ello.

Estrategias de Transfer Learning



Estrategias de Transfer Learning

Categorías

- **Transfer Learning Inductivo:**

- El dominio destino tiene datos etiquetados, el origen puede que sí o no.
- Cuando dominios origen y destino son el mismo, pero las tareas origen y destino son diferentes.

- **Transfer Learning No Supervisado:**

- No hay datos etiquetados en ambos dominios.
- Igual que inductivo, pero con foco en tareas no supervisadas en el dominio objetivo.

- **Transfer Learning Transductivo:**

- El dominio origen tiene muchos datos etiquetados, pero dominio destino no.
- Cuando similitudes entre tareas origen y destino, pero dominios diferentes.

Estrategias de Transfer Learning

¿Qué transferir?

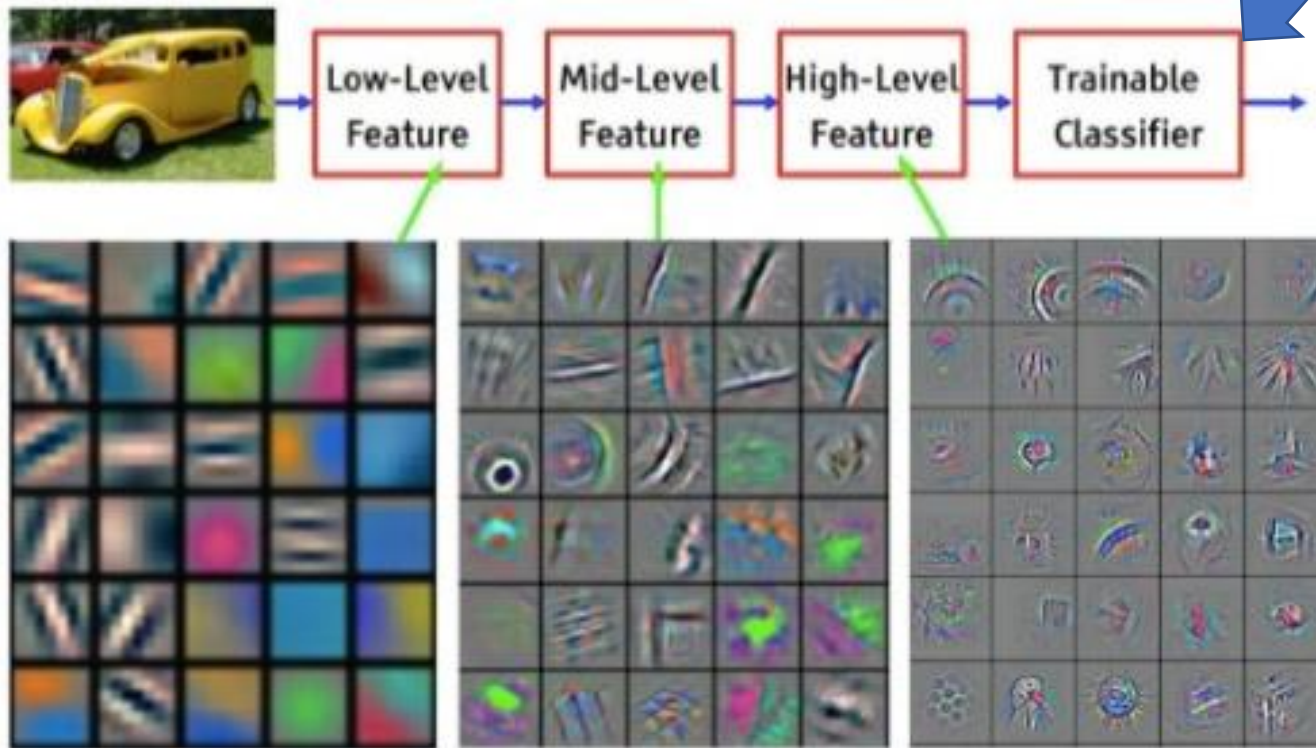
- **Transferencia de instancias:** Reusar ciertas instancias del dominio origen con los datos objetivo.
- **Transferencia de representación de características:** minimizar la divergencia de dominios identificando buenas representaciones de características que se puedan usar del dominio origen en el objetivo.
- **Transferencia de parámetros:** Cuando los modelos para tareas relacionadas comparten algunos parámetros o distribución de hiperparámetros.
- **Transferencia de conocimiento relacional:** Para manejar datos relacionados entre sí (como redes sociales).

Transfer Learning en Deep Learning

- Para entrenar una red convolucional (CNN) desde cero se necesita:
 - **Much(ísim)os datos** (*p.ej. ImageNet: 1,2 millones de imágenes, 1000 categorías*)
 - Gran **capacidad computacional** (*p.ej. [DGX v2](#) con 16 Tesla V100*).
 - **Tiempo** (*semanas a meses para entrenamiento*)
- En la realidad, pocos investigadores entrenan una CNN desde cero
 - Partir de una ConvNet pre-entrenada en un conjunto de datos muy grande
- Yosinski et al. [How transferable are features in deep neural networks?](#) 2014

Transfer Learning en Deep Learning

Convolutional Neural Network



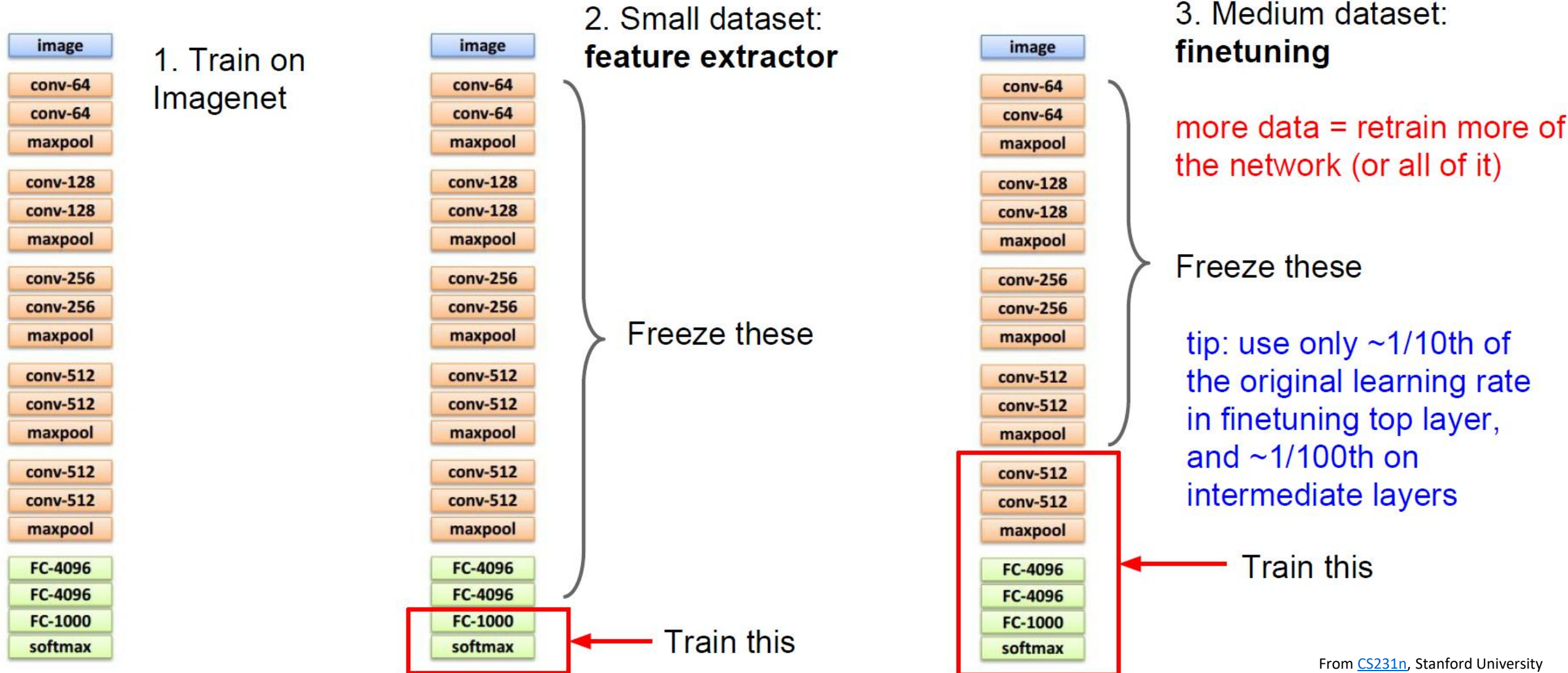
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Transfer Learning en Deep Learning

- Tres escenarios:
 - **Fixed feature extractor:** eliminar las últimas capas FC (Fully Connected) y el clasificador, y fijar el resto.
 - Re-entrenar el clasificador con el nuevo conjunto de datos.
 - **Fine-tuning:** fijar solo las primeras capas, aplicar *backpropagation* al resto.
 - Últimas capas suelen contener características más específicas a las categorías por las que fueron entrenadas.
 - **Pretrained models:** descargar un modelo (p.ej. [Model Zoo de Caffe](#)) para aplicar lo anterior.

<http://cs231n.github.io/transfer-learning/>

Transfer Learning en Deep Learning



Transfer Learning en Deep Learning

	very similar dataset	very different dataset
very little data	Use Linear Classifier on top layer	You're in trouble... Try linear classifier from different stages
quite a lot of data	Finetune a few layers	Finetune a larger number of layers

Transfer Learning en Deep Learning

- Aplicaciones:

- Para visión por computador: Quizás donde más se esté aplicando esta técnica hoy en día.
 - Modelos pre-entrenados: [VGG-16](#), [VGG-19](#), [Inception V3](#), [Xception](#), [ResNet-50](#)
- Para procesamiento de lenguaje natural: Aquí se hace difícil pero se puede reutilizar modelos de word embedding.
 - Modelos pre-entrenados: [Word2Vec](#), [GloVe](#), [FastText](#)
- Para audio/habla: modelos de reconocimiento automático del habla (ASR) desarrollados para el inglés se han usado con éxito para otros lenguajes como el alemán.

Agenda

- Conceptos básicos
- **Casos de estudio:**
 - **Tag extraction**
 - Glaucoma
- Tutorial con Keras:
 - Knifey-spoony
 - Sesión 1
 - Sesión 2
 - Ejercicios

Tag extraction

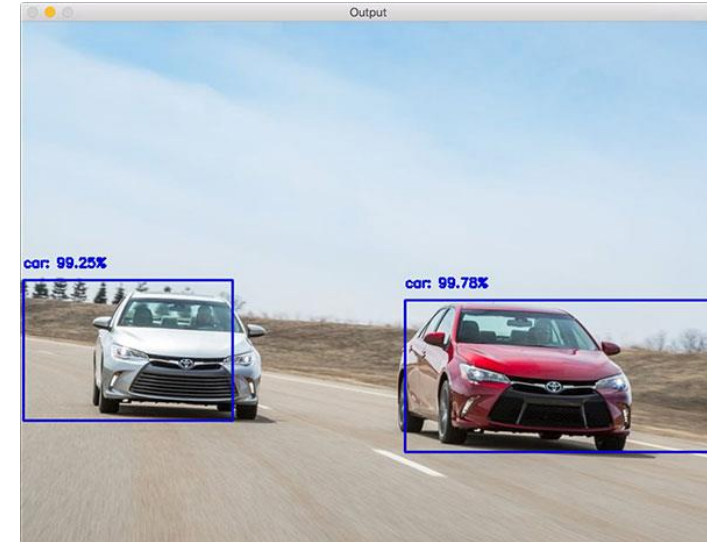
Introducción

- Trabajo publicado en:
 - **U.A. Kahn**, N. Ejaz, M.A. Martínez-del-Amor, H. Sparenberg. [Movies Tags Extraction Using Deep Learning](#). *14th IEEE International Conference on Advanced Video and Signal Surveillance (AVSS 2017), Lecce, Italy, 29 August - 1 September 2017*. Proceedings (October 2017), IEEE Xplore, pp. 1-6.
- Enlace al proyecto:
 - <http://umair-khan.quest.edu.pk/announcements>

Tag extraction

(extracción de etiquetas)

- **Extracción** (automática) de etiquetas de películas:
 - *Metadatos imprecisos* generado por humanos
 - Extracción de *información saliente* usando machine learning
 - *Semántica* de “alto nivel”
- **Idea:** selección de etiquetas clave, representando el tema general
- **Diferente** a la mayoría de tareas de reconocimiento de objetos o escenas:
 - No nos interesa si en un vídeo aparece una pistola
 - Nos interesa saber si el video es de violencia, acción, etc.



Object detection: 2 cars

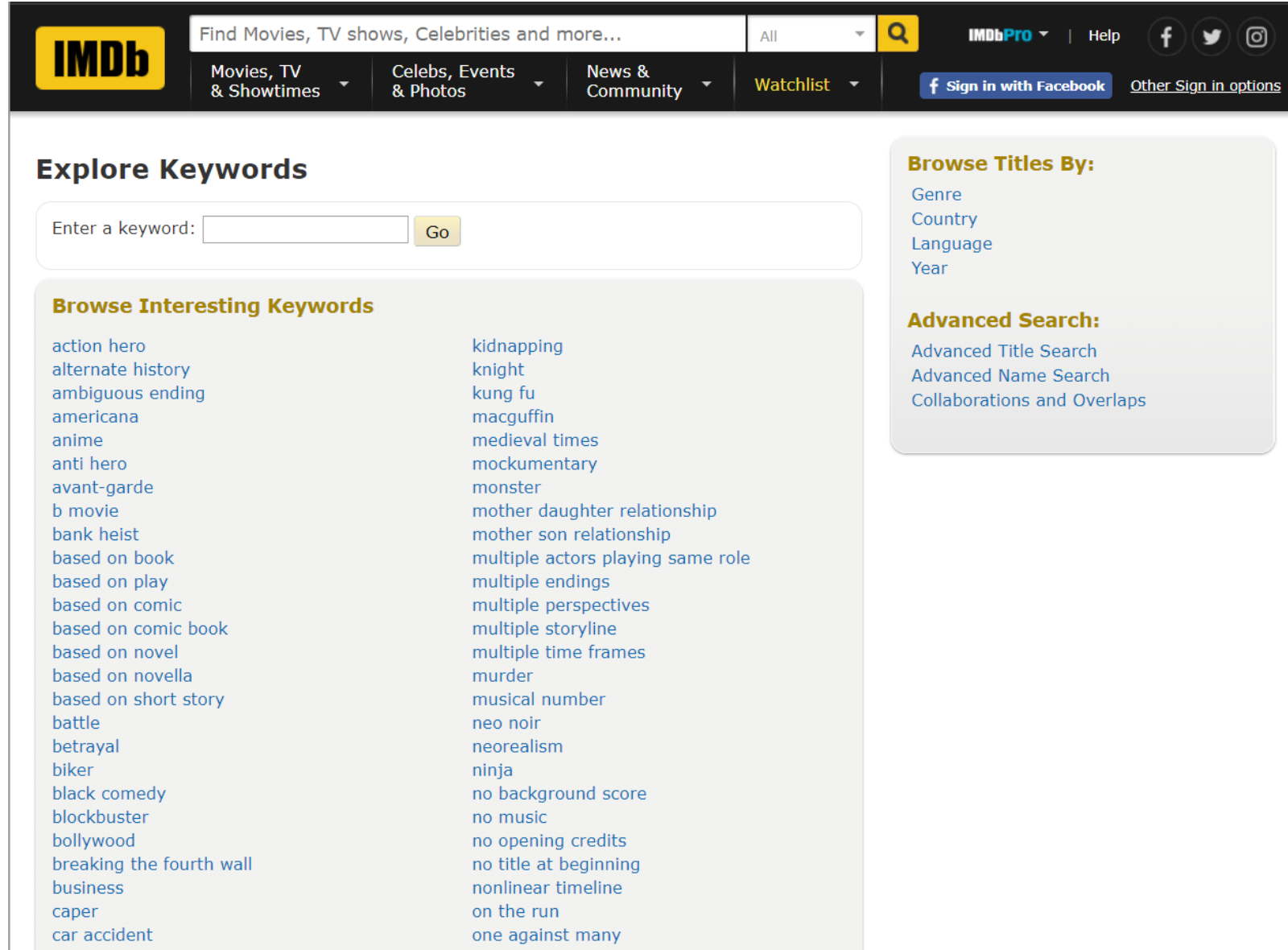
vs

Movie tag: car chase





Tag extraction

Introducción

- Etiquetas \approx keywords de IMDb
- Haremos solo una selección de ellas



The screenshot shows the IMDb website interface. At the top is a navigation bar with the IMDb logo, a search bar, and various links. Below the navigation bar, the 'Explore Keywords' section is visible, featuring a search input field and a 'Go' button. To the right of the search input is a 'Browse Interesting Keywords' section with two columns of keyword tags. Further to the right is a 'Browse Titles By:' section with links for Genre, Country, Language, and Year. Below that is an 'Advanced Search:' section with links for Advanced Title Search, Advanced Name Search, and Collaborations and Overlaps.

IMDb Find Movies, TV shows, Celebrities and more... All  **IMDbPro** | Help   

Explore Keywords

Enter a keyword: **Go**

Browse Interesting Keywords

action hero	kidnapping
alternate history	knight
ambiguous ending	kung fu
americana	macguffin
anime	medieval times
anti hero	mockumentary
avant-garde	monster
b movie	mother daughter relationship
bank heist	mother son relationship
based on book	multiple actors playing same role
based on play	multiple endings
based on comic	multiple perspectives
based on comic book	multiple storyline
based on novel	multiple time frames
based on novella	murder
based on short story	musical number
battle	neo noir
betrayal	neorealism
biker	ninja
black comedy	no background score
blockbuster	no music
bollywood	no opening credits
breaking the fourth wall	no title at beginning
business	nonlinear timeline
caper	on the run
car accident	one against many

Browse Titles By:

- [Genre](#)
- [Country](#)
- [Language](#)
- [Year](#)

Advanced Search:

- [Advanced Title Search](#)
- [Advanced Name Search](#)
- [Collaborations and Overlaps](#)

Tag extraction

Introducción

- Aplicaciones potenciales:
 - Búsqueda por consulta
 - Almacenamiento y clasificación eficiente
 - Censura de contenido (violencia, desnudez, etc)
 - Sistemas de recomendación
 - Recuperación guiada por la escena
 - Asistencia en traducción de películas a lenguaje natural
 - Reconocimiento de acciones y comportamiento

Tag extraction

Diseño conceptual

- **Problemática:**

- No existe un dataset para etiquetas
- Confección de uno desde cero y de forma manual
- No disponíamos de recursos computacionales ni de mucho tiempo (beca posdoctoral).

- **Solución:** Transfer Learning para Fixed Feature Extractor

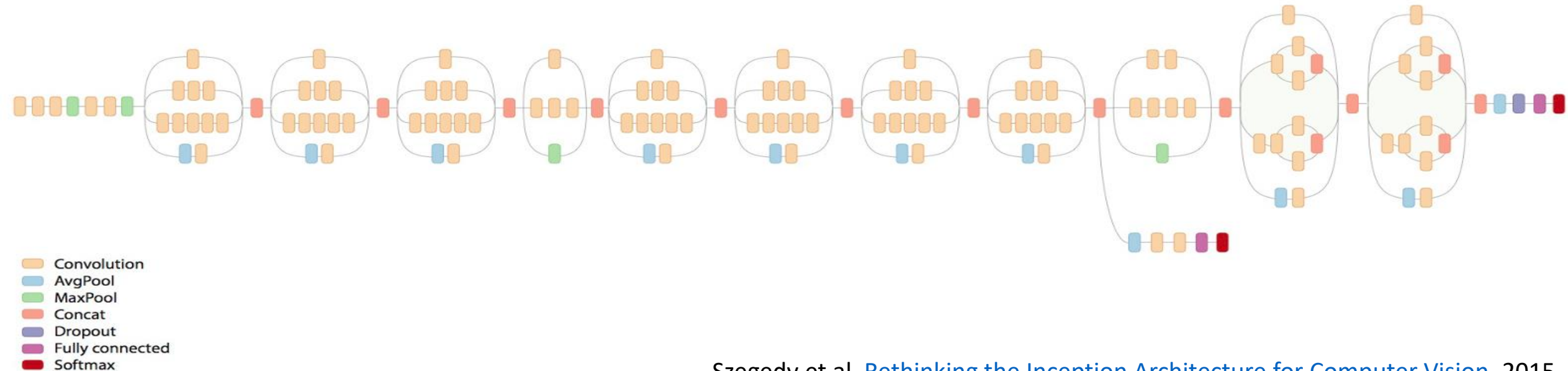
- Necesitamos un dataset “mediano”
- Inception-v3: Suficiente para nuestros PCs

Tag extraction

Elección modelo pre-entrenado

- **Inception-v3:**

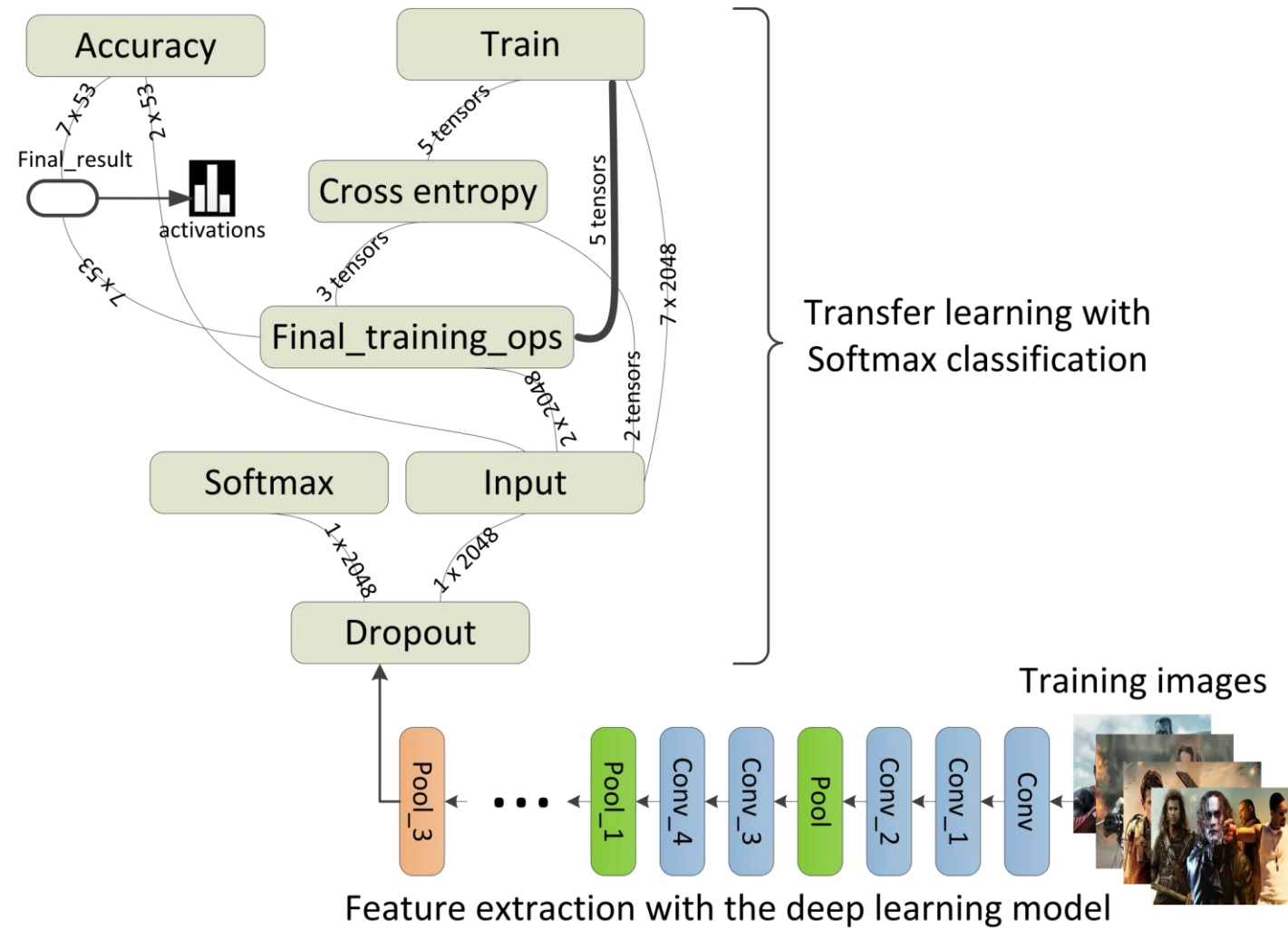
- Entrenado sobre ImageNet 2012 y 1000 clases
- 3.46% tasa error
- 48 capas
- Disponible en tensorflow y [tutoriales](#)



Tag extraction

Transfer learning sobre Inception-v3

- Eliminada última capa
- Añadida una capa **Dropout**:
 - Desecha 50% activaciones aleatoriamente
 - Evita overfitting
- Después, activación **ReLU**:
 - No linealidad
 - $y_i = \text{ReLU}[\sum_j W_{i,j}x_j + b_i]$
- Clasificación con **Softmax**:
 - Convertir a probabilidad
 - $$p_i = \frac{e^{y_i}}{\sum_j^{50} e^{y_j}}$$



Tag extraction

Conjunto de datos

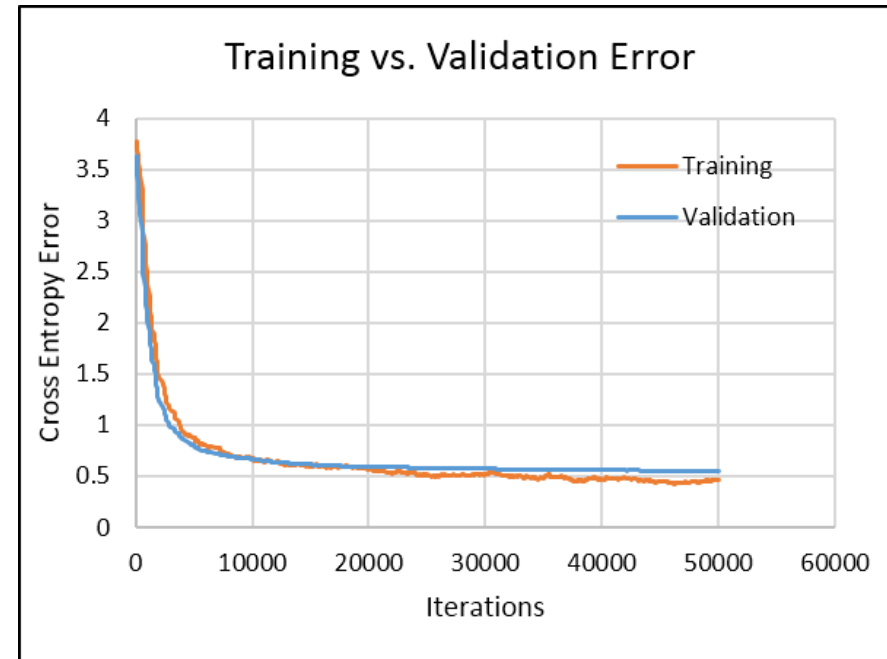
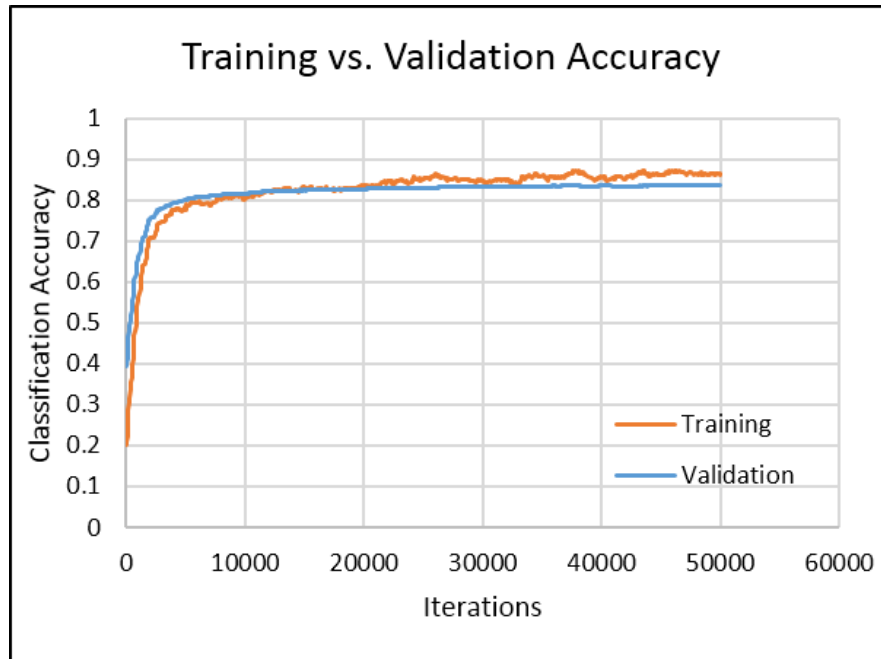
- Vocabulario 50 etiquetas (con solapamiento)
- 700 imágenes/etiqueta

Action	Bomb explosion	Car chase
Destruction	Sword fight	Vehicle crash
Violence	Abduction	Heist
Adventure	Animal	Beach/Sea
Climbing	Desert	Hiking
Forest	Valleys/Hills	Children
Family	Club/Bar	Dance
Music	Wedding	College/Univ.
Hospital	Drinking	Food
Smoking	Exercise	Sports
Swimming	Glamor/Fashion	Nudity
Romance	Sex	Horror
Monster	Murder	Lab Experiment
Sci-fi	Super hero	Technology
Robot	Military	Police
Prison	War	Weapon
Animation	Drama	

Tag extraction

Resultados en fotogramas individuales

- **Distribución:** 80% entrenamiento, 10% validación y 10% test
- **Cross entropy**, función de pérdida para estimar error: $E(p, q) = - \sum_x^{50} q(x) \log p(x)$
- **500 epochs**



Tag extraction

Resultados en fotogramas individuales



Military, action, weapon, war



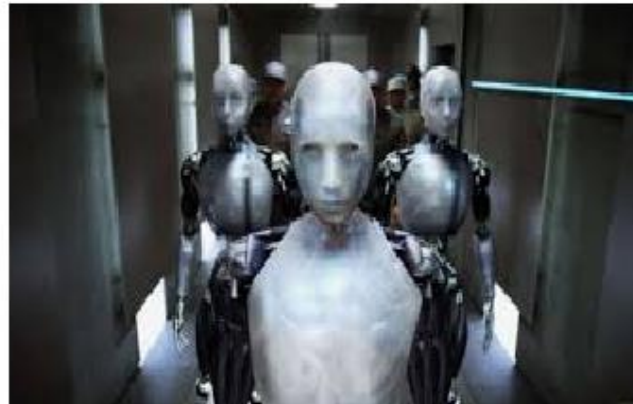
Violence, destruction, bomb explosion, action, car crash



Sex, nudity, romance, modeling



Hiking, adventure, nature, forest, valleys, hills, climbing



Sci-fi, super hero, robot, action

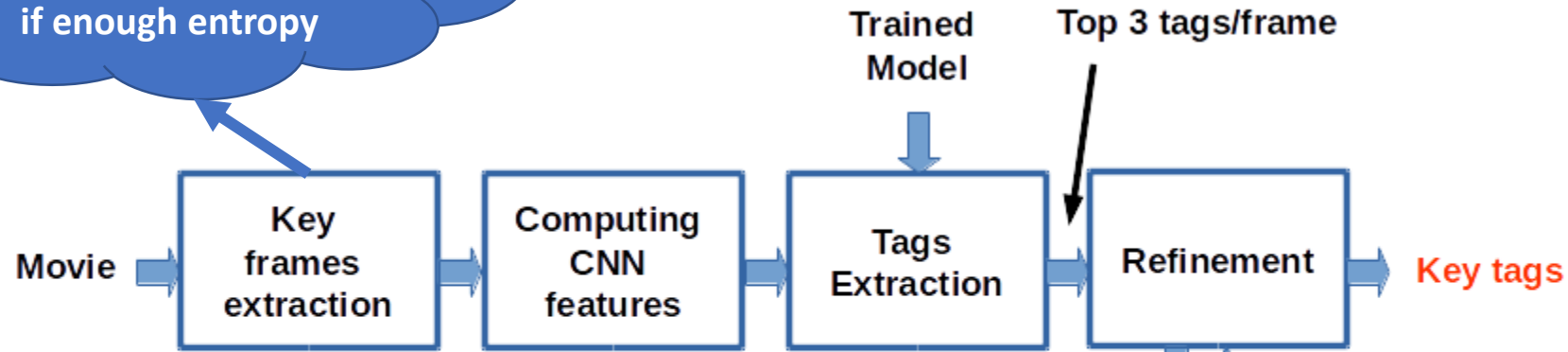


Violence, sci-fi, action, horror

Tag extraction

Extrapolación a vídeo

- Finding shot boundary
- Middle frame of shot, if enough entropy



$$W_i = \frac{n_i}{N} \sum_{j=0}^N P_{ij}$$

$$R_i = \frac{W_i - W_{min}}{W_{max} - W_{min}}$$

Tag extraction

Experimentación

- Configuración Hardware/Software:

Hardware/Software	Specifications
CPU	Intel Xeon(R) E5430, 2.66GHz x 8
RAM	8GB
GPU	GeForce GTX 1050 Ti, 768 cores, 4GB GDDR5
Deep learning framework	Tensorflow 1.0, compiled with GPU support
Operating System	Ubuntu 16.04 (64-bit)
Programming languages	Python 2.7, OpenCV 3.0, C++

Tag extraction

Experimentación

- Problemática:
 - **No hay un ground truth**, o marco de referencia
- Realización de **3 experimentos subjetivos**:
 - Llevados a cabo en el cine del Fraunhofer IIS
 - Muestra de 10 tráilers de películas
 - 10 voluntarios distintos en cada uno
- **Medidas**:
 - Mean Opinion Score: *de una encuesta, cuantos resultados corresponden*
 - Precisión: *cuántos de los resultados positivos son correctos*
 - Recall: *cuántos resultados positivos correctos respecto a todos los positivos*
 - F-score: *media ponderada de precisión y recall*



Tag extraction

Experimentación

- **Experimento 1: tags rating**

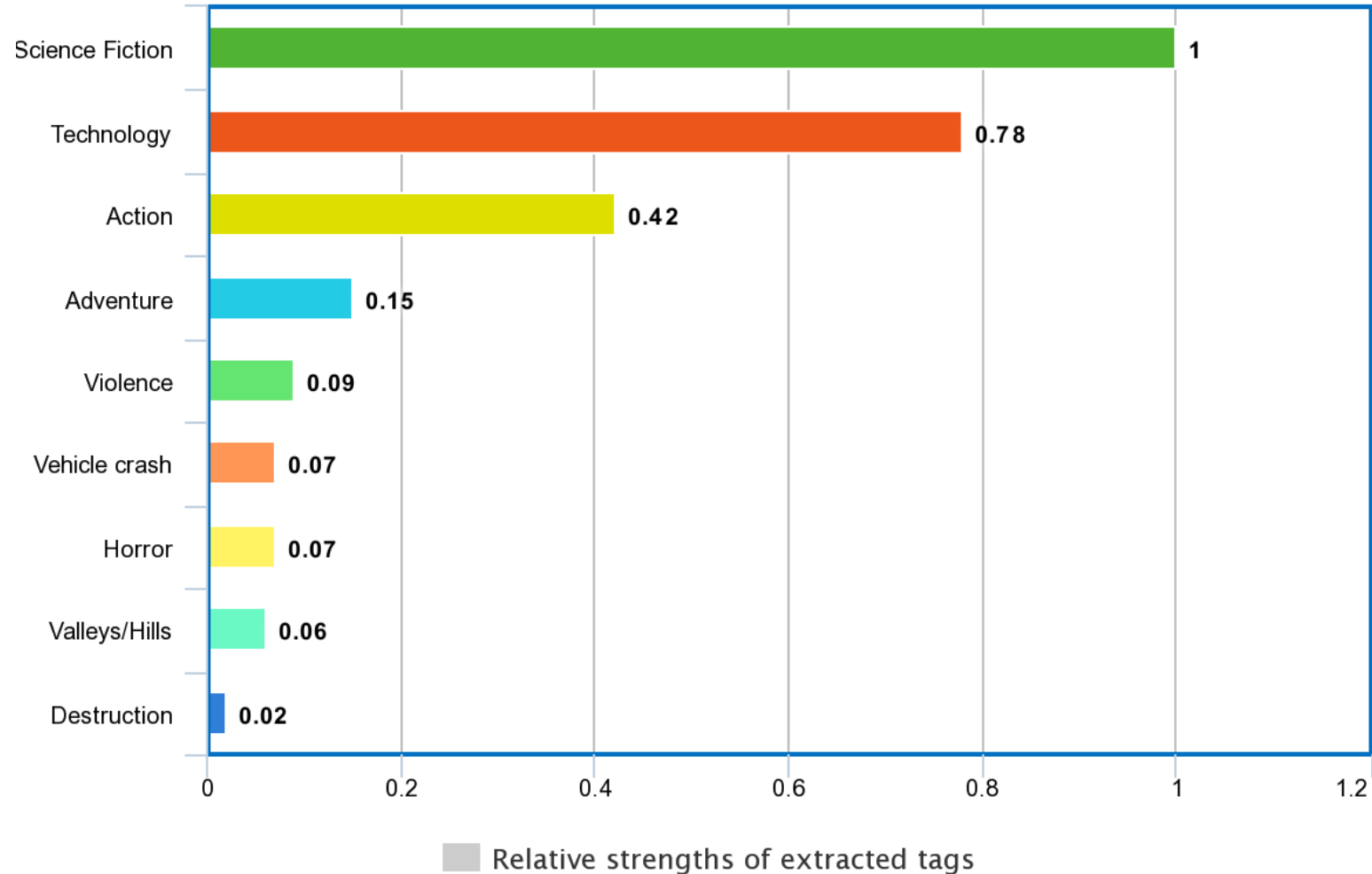
- Reparto de las etiquetas extraídas por nuestro algoritmo para cada vídeo
- Voluntarios valora cada uno por separado.
- Mean Opinion Score: 84.3%
- Ejemplo: [RAW \(2017\) tráiler](#)
 - Romance, violence, action, car crash, horror, sex, child, nudity, outdoor/nature/forest, hospital, food, Club/bar, college/university, music, crowd

Tag extraction

Experimentación

- **Experimento 2: tags rating w.r.t. relevancy and strength**

- Igual que experimento 1, pero teniendo en cuenta la relevancia de cada etiqueta.
- Mean Opinion Score: **77.8%**
- Ejemplo: [Alien Covenant teaser trailer](#)



Tag extraction

Experimentación

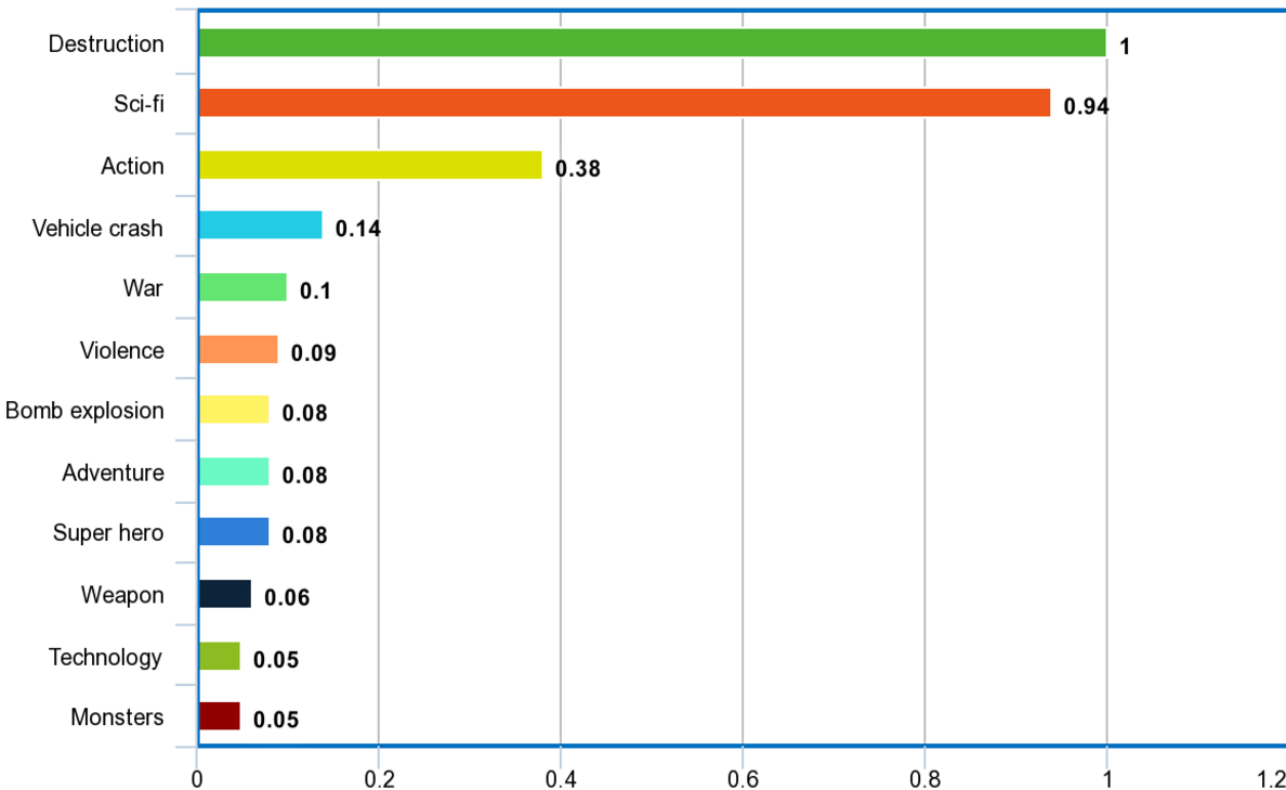
- **Experimento 3: tags matching**

- Repartir todo el vocabulario de etiquetas, repetido para cada vídeo
- Voluntarios eligen las etiquetas que crean más relevantes
- Usando este experimento como ground truth:
 - MAP = 76%, MAR = 74.22%
 - F1 - Score = 0.75%

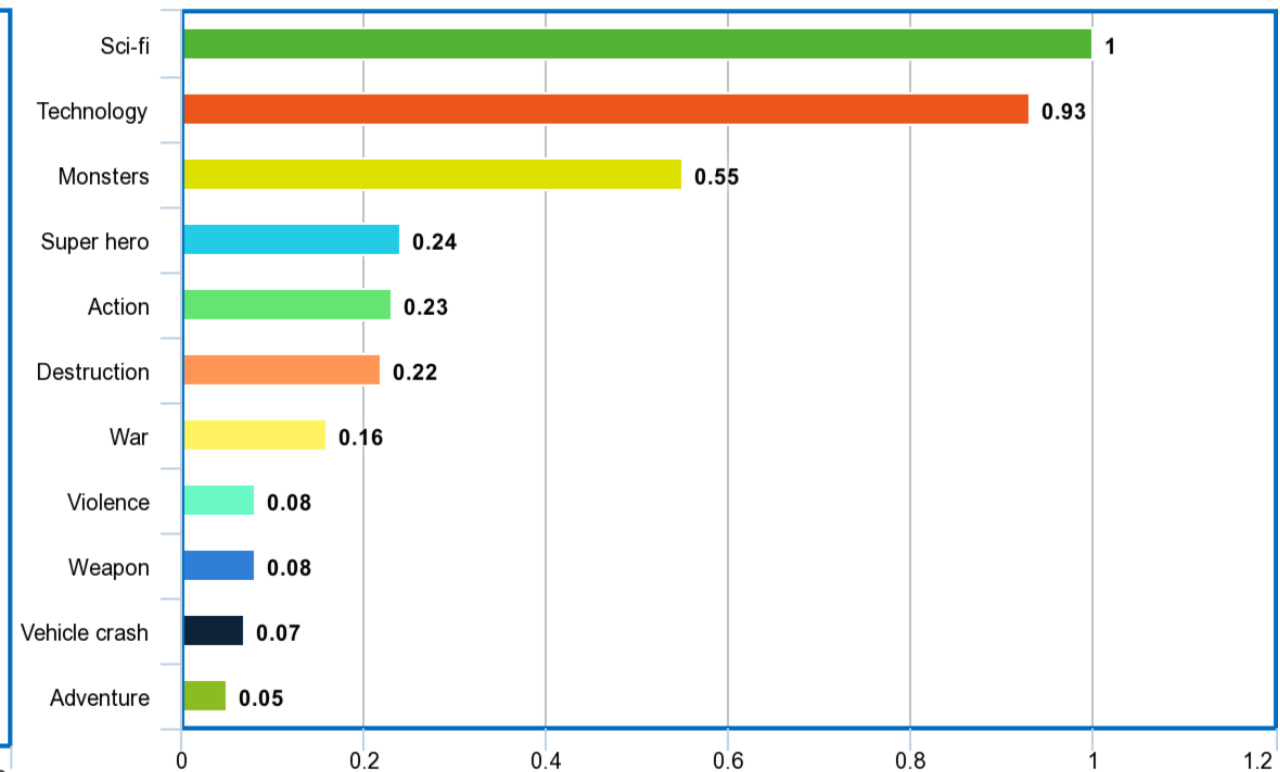
Tag extraction

Experimentación ([tráiler](#) vs película completa)

- **Trailer duration:** 2 min, 26 sec, **Processing time:** 17sec
- **Full length movie duration:** 1 hr, 24 min, **Processing time:** 10 min



■ The Guardians (2017) Trailer

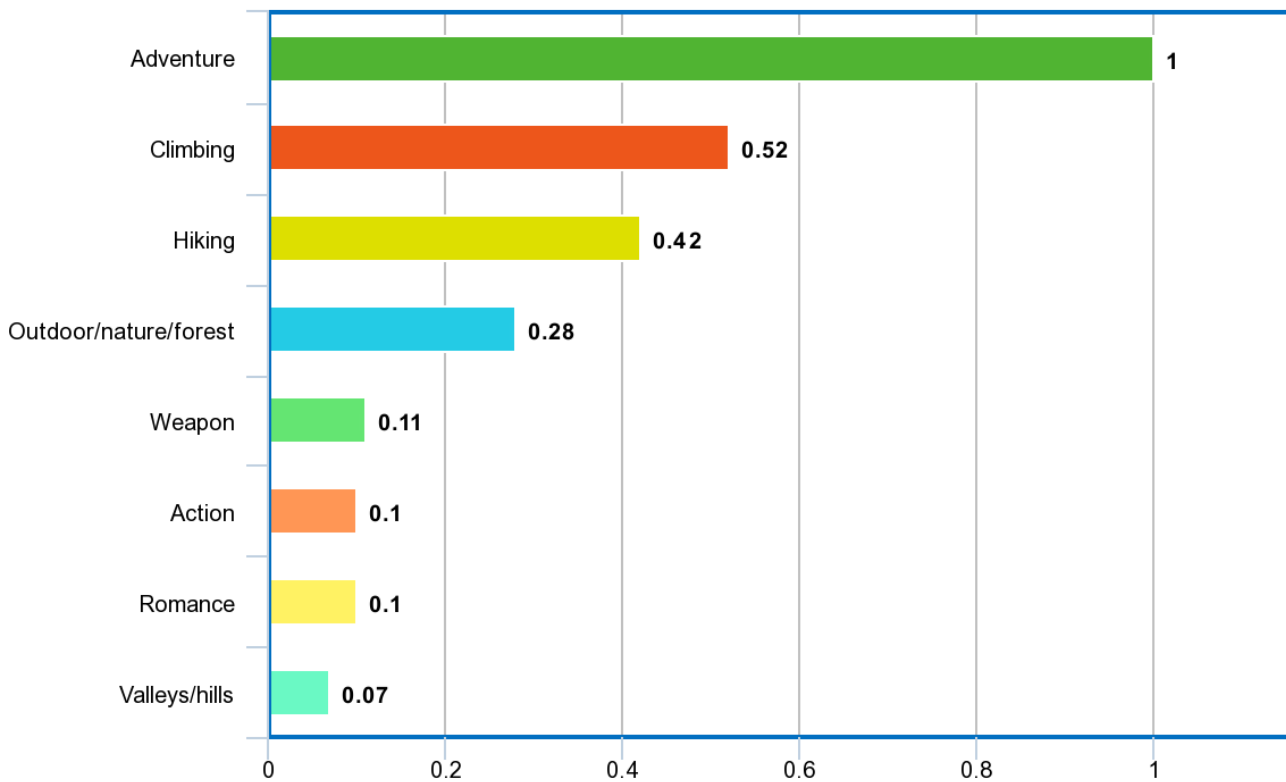


■ The Guardians Full Length Movie

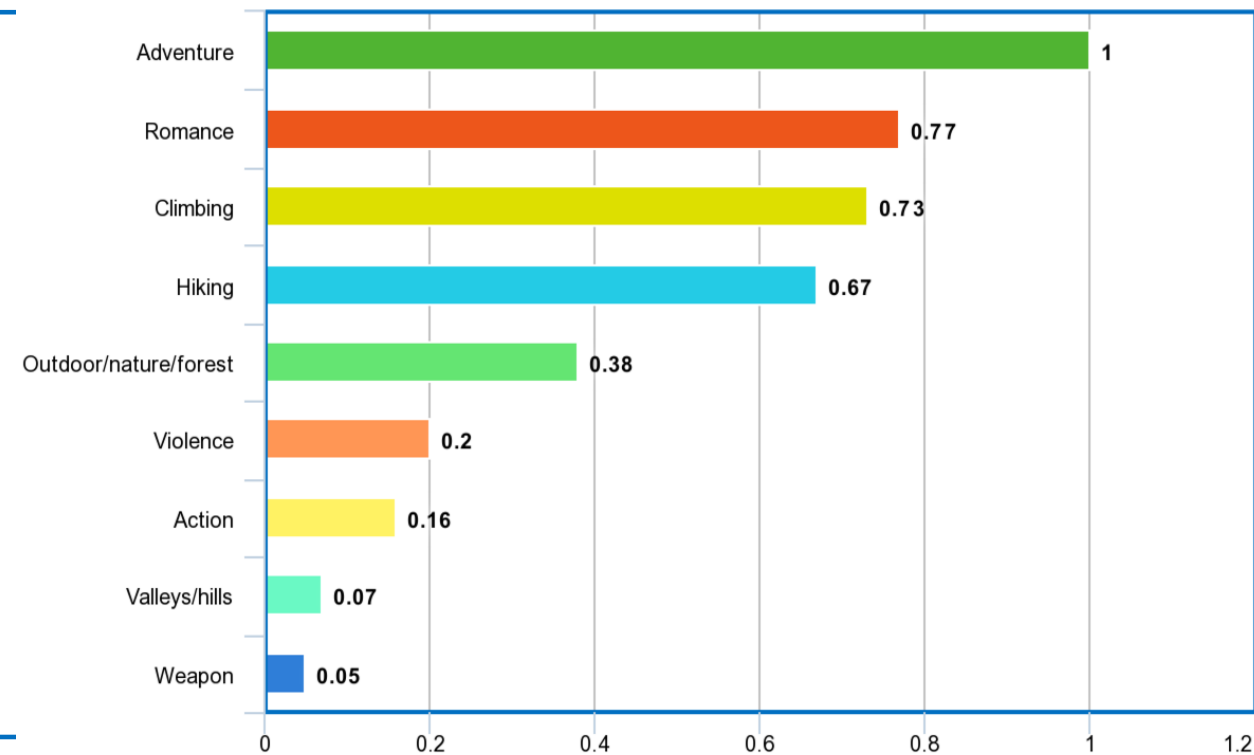
Tag extraction

Experimentación (tráiler vs película completa)

- **Trailer duration:** 1 min, 53 sec, **Processing time:** 13 sec
- **Full length movie duration:** 1 hr, 39 min, **Processing time:** 7 min



■ A lonely place to die trailer



■ A lonely place to die (Full length)

Agenda

- Conceptos básicos
- Casos de estudio:
 - Tag extraction
 - **Glaucoma**
- Tutorial con Keras:
 - Sesión 1
 - Sesión 2
 - Ejercicios

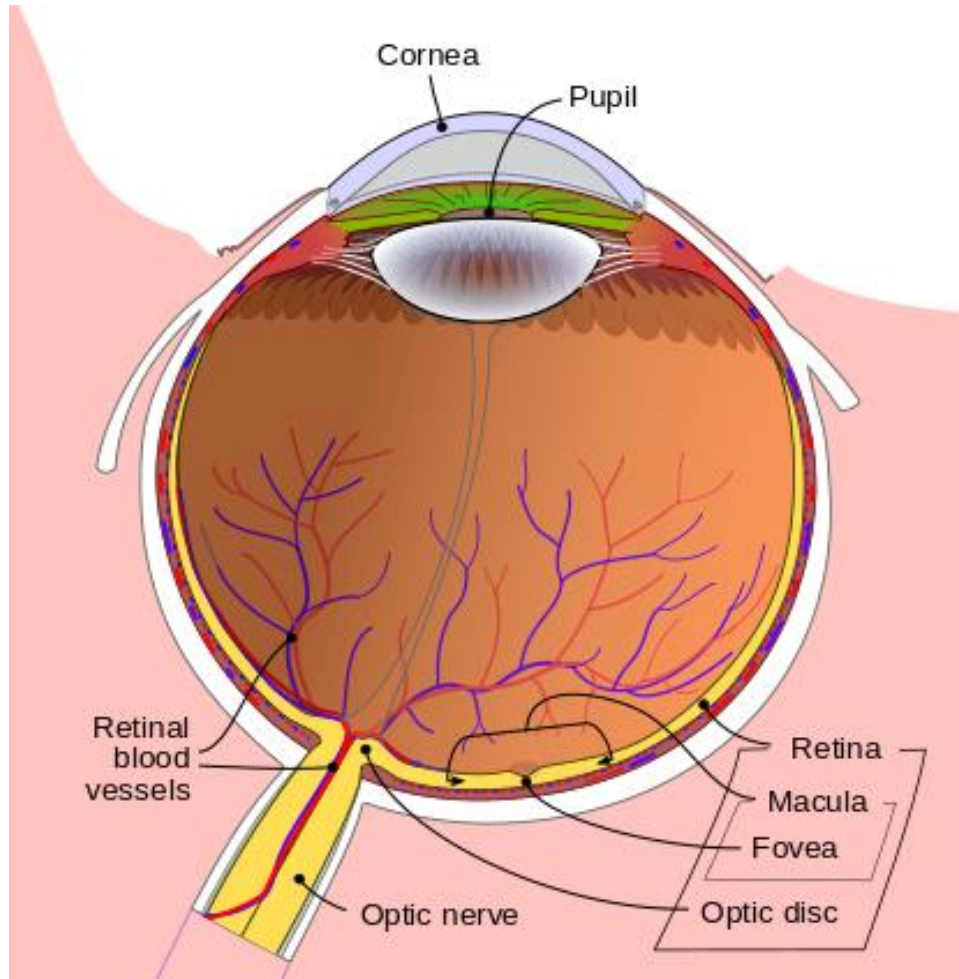
Glaucoma

Clasificación del fondo de ojo

- Trabajo publicado en:
 - **Diaz-Pinto A**, Morales S, Naranjo V, Köhler T, Mossi J M, Navea A. [CNNs for Automatic Glaucoma Assessment using Fundus Images: An Extensive Validation](#). BioMedical Engineering OnLine 2019 18:29, doi:10.1186/s12938-019-0649-y. 2019
 - Diapositivas completas: [seminario \(I+A\)A](#)

Glaucoma

Clasificación del fondo de ojo



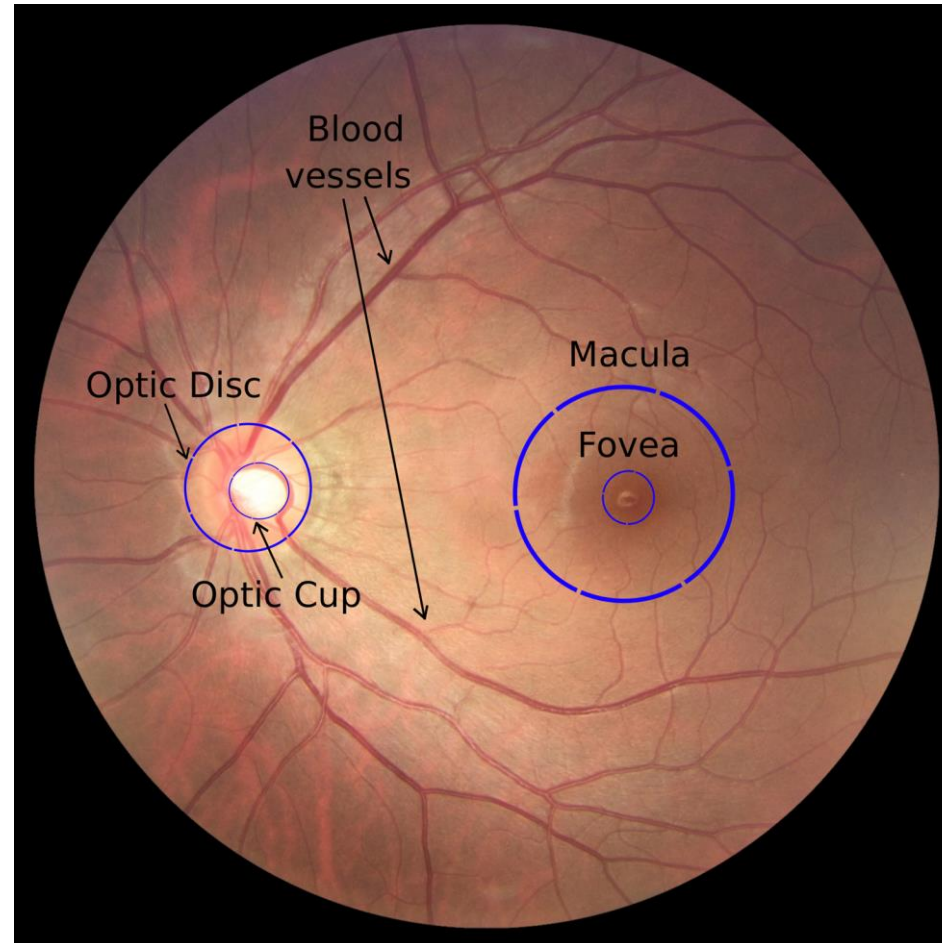
Anatomía de la retina.

Tres estructuras principales:

- El disco óptico
- Vasos sanguíneos retinales
- La macula

Glaucoma

Clasificación del fondo de ojo



Glaucoma

Tipos

- El glaucoma se refiere a la profundización o excavación de la cabeza del nervio óptico.
- Y hay tres formas principales de glaucoma:
 - Glaucoma de ángulo abierto
 - Glaucoma de cierre de ángulo y
 - Glaucoma congénito

Glaucoma

ImageNet-trained CNN architectures

- Arquitecturas CNN entrenadas por ImageNet aplicadas para la clasificación de imágenes retinianas:
 - **VGG16** y **VGG19**: Estas CNNs se basan en el mismo modelo y se caracterizan por su simplicidad. Presentado por Simonyan en 2014 para el reto ImageNet

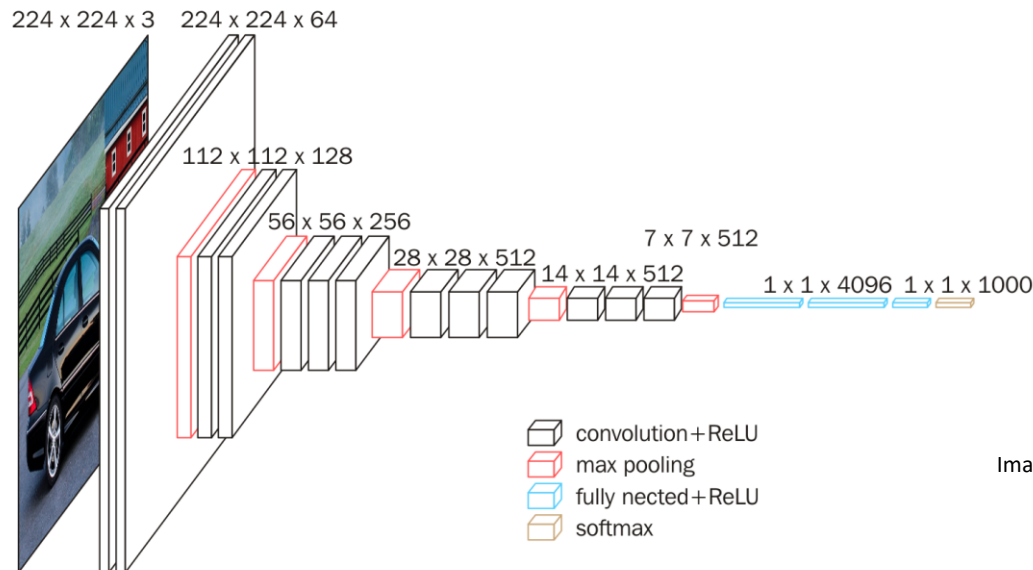
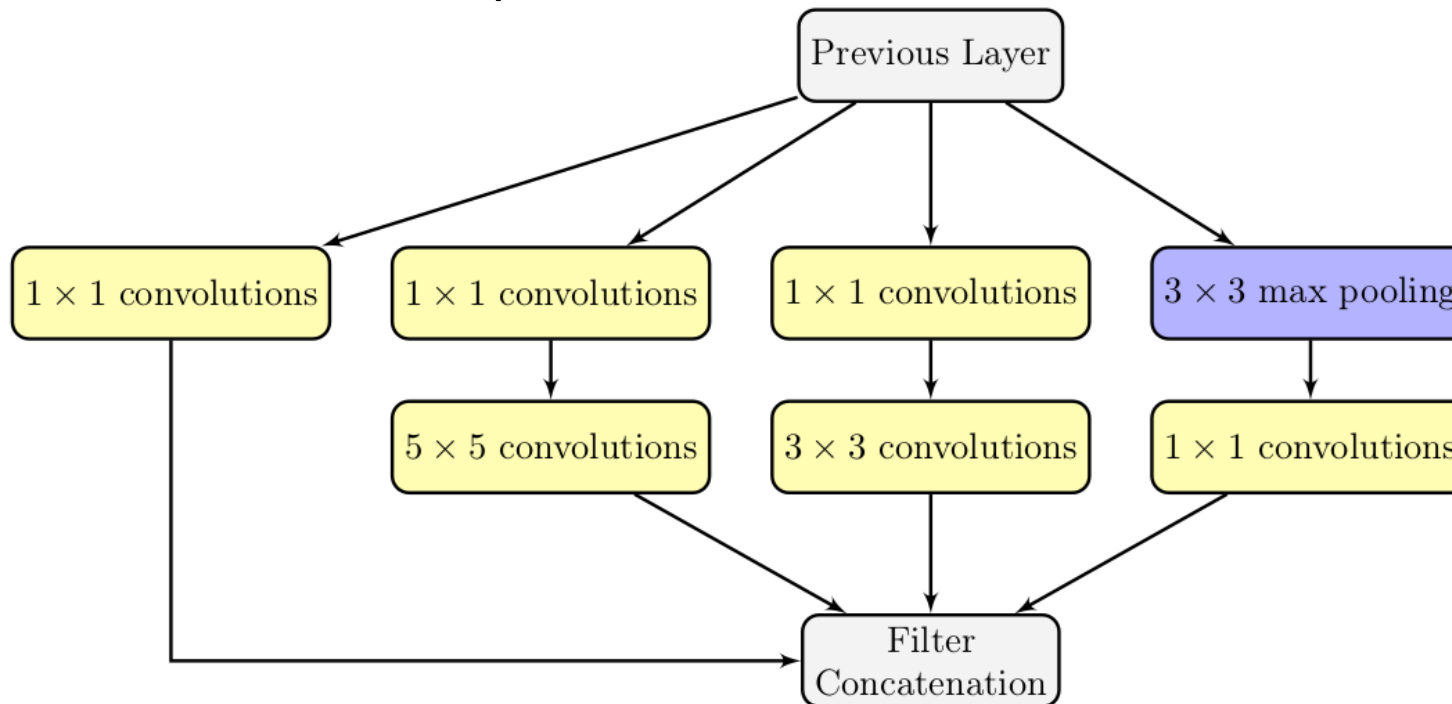


Image downloaded from: <https://blog.datawow.io/cnn-models-ef356bc11032>

Glaucoma

ImageNet-trained CNN architectures

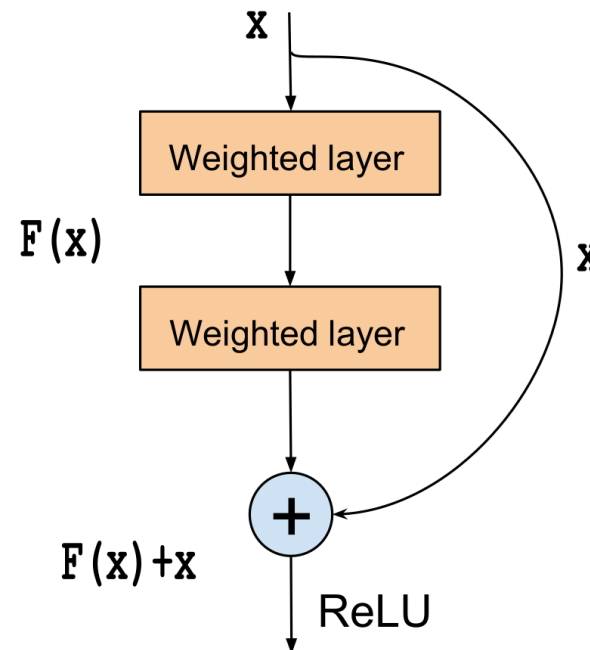
- Arquitecturas CNN entrenadas por ImageNet aplicadas para la clasificación de imágenes retinianas:
 - **GoogLeNet**: Fue introducido por primera vez por Szegedy et al. en 2015. Se basa en el módulo Inception.



Glaucoma

ImageNet-trained CNN architectures

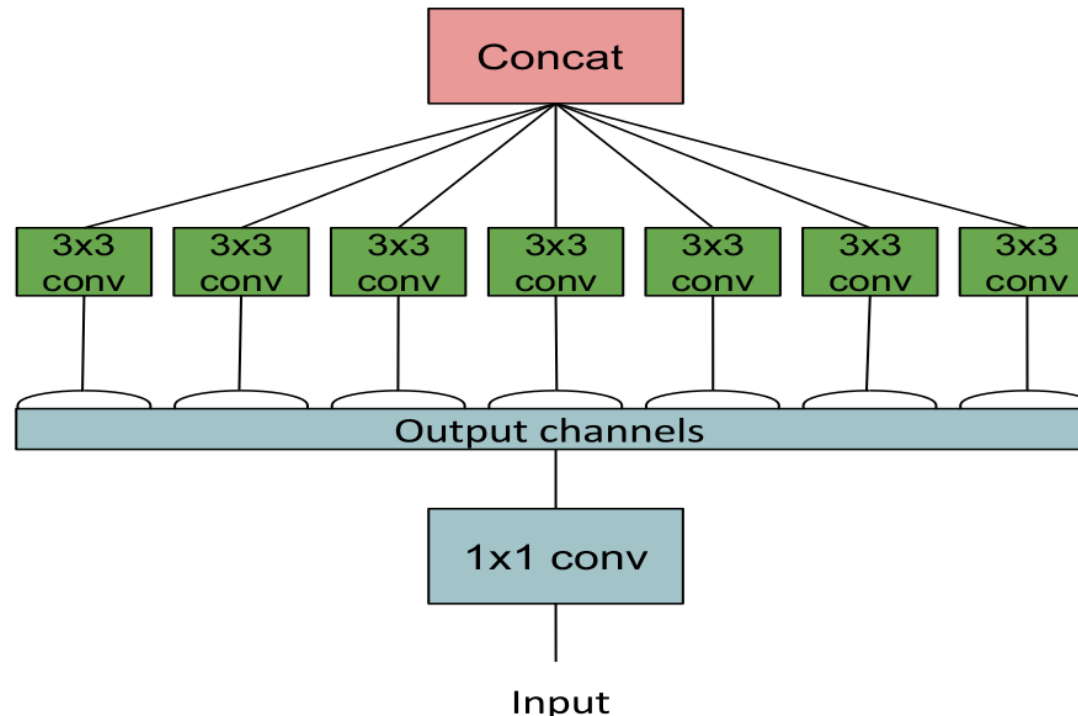
- Arquitecturas CNN entrenadas por ImageNet aplicadas para la clasificación de imágenes retinianas:
 - Microsoft **ResNet**: Esta arquitectura fue propuesta por el equipo de Microsoft Research Asia (MSRA) en 2015. Es considerada una "arquitectura exótica" que se basa en bloques residuales.



Glaucoma

ImageNet-trained CNN architectures

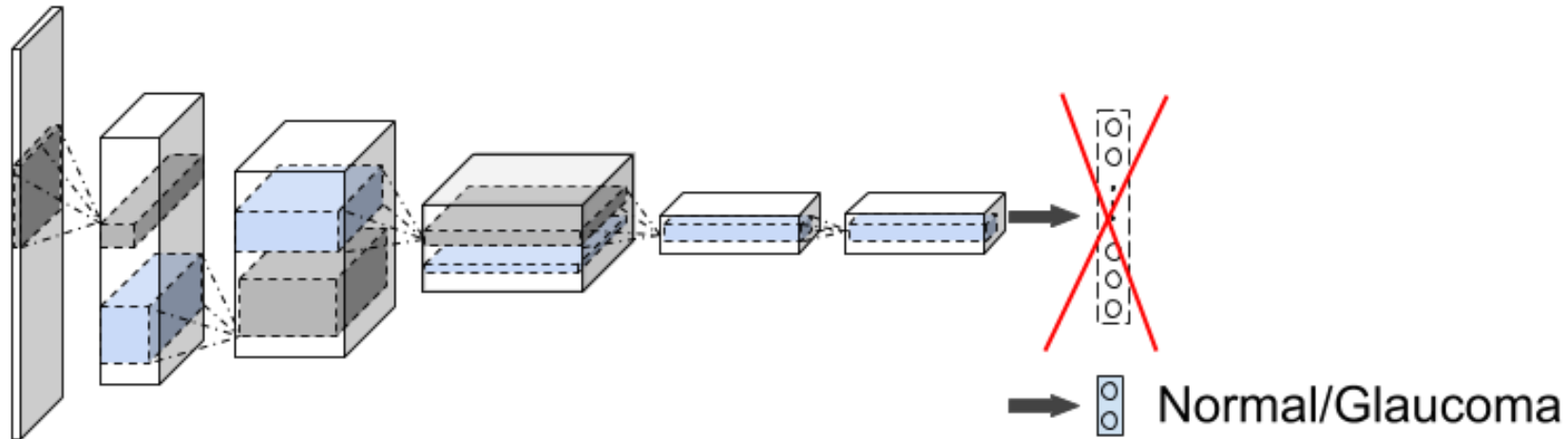
- Arquitecturas CNN entrenadas por ImageNet aplicadas para la clasificación de imágenes retinianas:
 - **Xception**: o Extreme Inception, fue propuesto por F. Chollet en 2016. Es una extensión de la arquitectura de Inception.



Glaucoma

ImageNet-trained CNN architectures

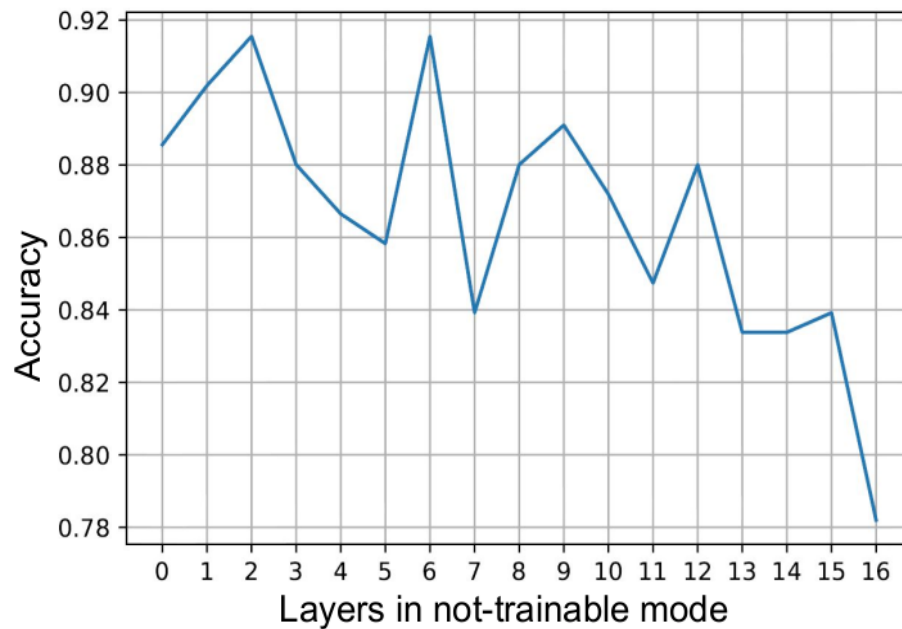
- Técnica de ajuste fino:
 - a) Inicialización del peso de las capas convolucionales utilizando los pesos de ImageNet y
 - b) Sustitución de la función de clasificación o del número de nodos en la última capa completamente conectada.



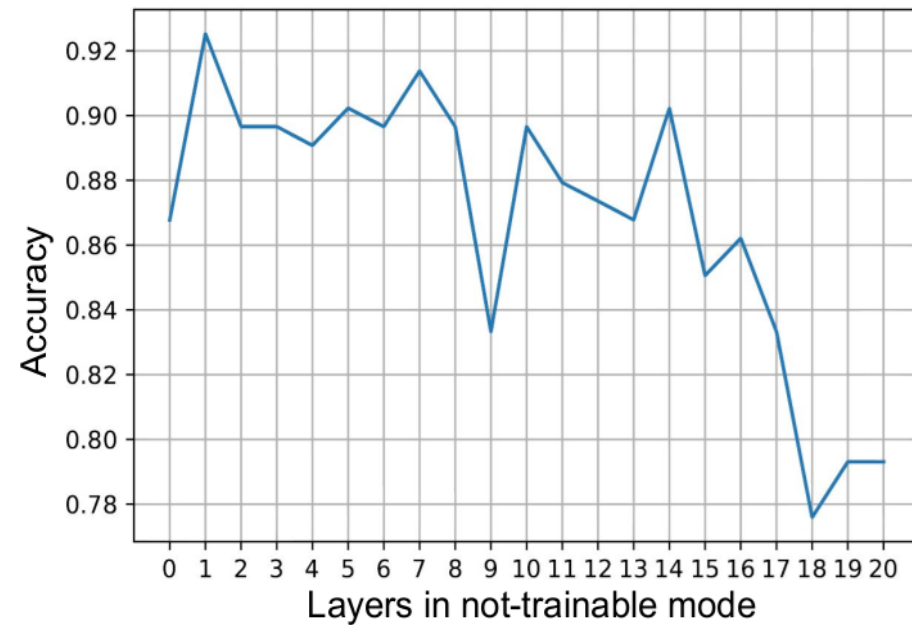
Glaucoma

ImageNet-trained CNN architectures

- Ajuste profundo (Deep) o superficial (shallow):



VGG16



VGG19

Glaucoma

ImageNet-trained CNN architectures

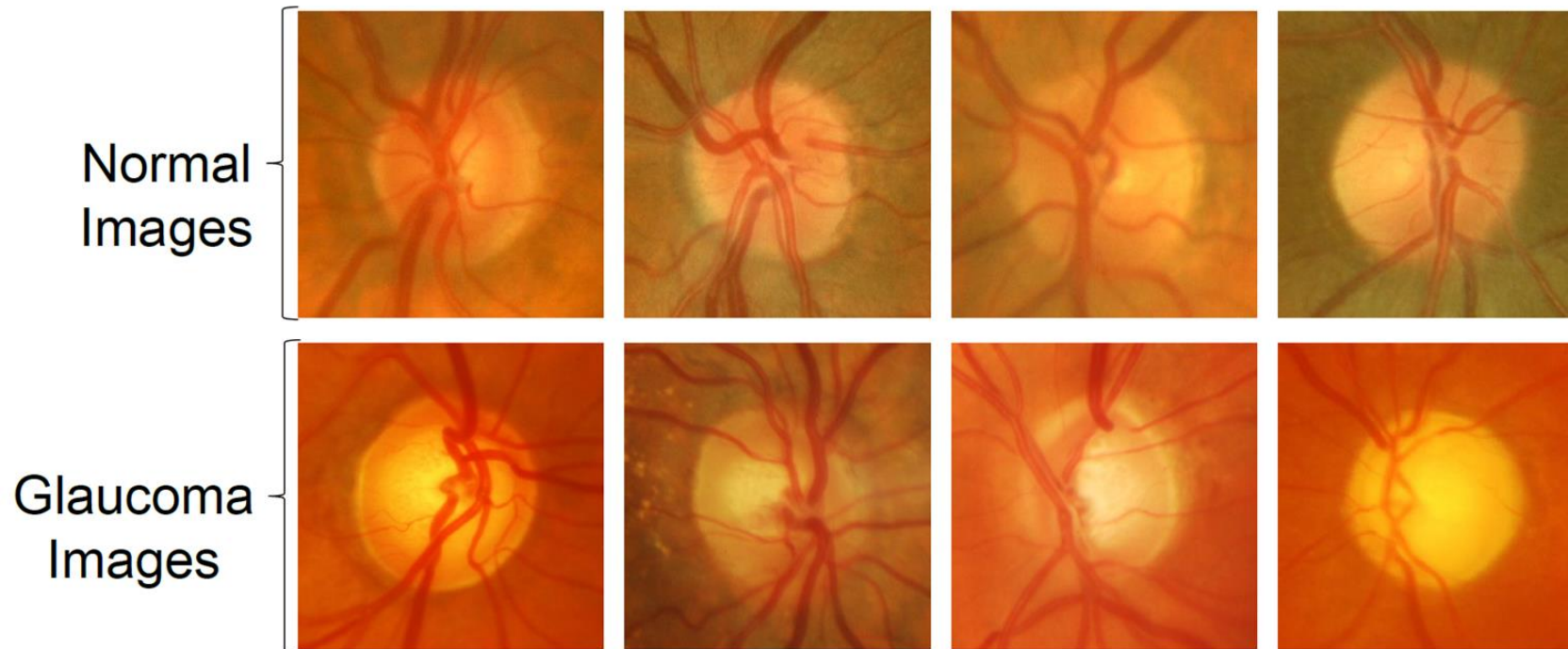
- Bases de datos:

Name	Glaucoma	Normal	Total
HRF	27	18	45
Drishti-GS1	70	31	101
RIM-ONE	194	261	455
sjchoi86-HRF	101	300	401
ACRIMA	396	309	705
Total	788	919	1707

Glaucoma

ImageNet-trained CNN architectures

- Todas estas imágenes se recortaron automáticamente alrededor del disco óptico utilizando un método de aprendizaje profundo¹:



1) Xu P, Wan C, Cheng J, Niu D, Liu J. Optic disc detection via deep learning in fundus images. Fetal, infant and ophthalmic medical image analysis.

Glaucoma

ImageNet-trained CNN architectures

- Las imágenes fueron reescaladas:
 - 224x224 px → VGG16, VGG19 y ResNet50
 - 299x299 px → InceptionV3 y Xception
- También se utilizó aumento de datos:
 - rotaciones de imágenes,
 - reflejo de imágenes,
 - deformación,
 - volteos verticales y horizontales.
- Validación cruzada de 10-fold
- Configuración de pruebas cruzadas

Glaucoma

ImageNet-trained CNN architectures


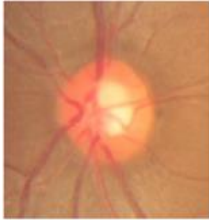

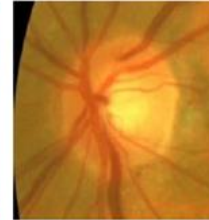
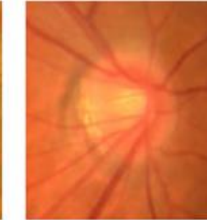
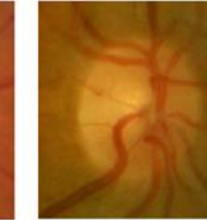
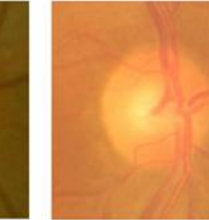
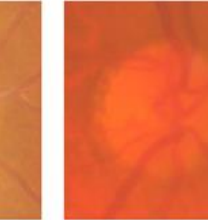
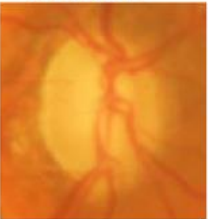
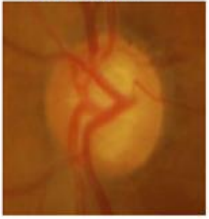
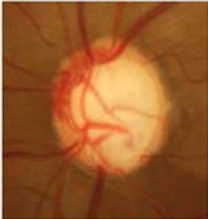
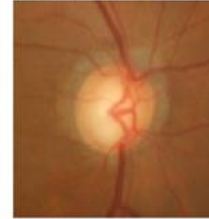
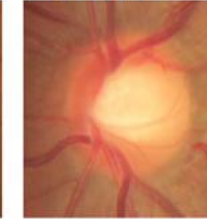
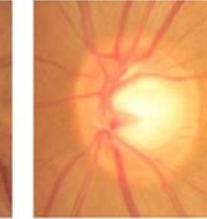
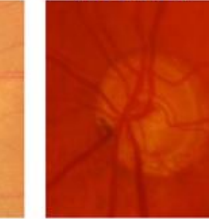
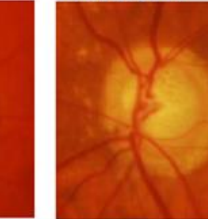
- Resultados validación cruzada de 10-fold

Model Name	AUC	Accuracy	F-score	# parameters (in millions)
VGG16	0.9632	0.8948	0.9005	138
VGG19	0.9686	0.9069	0.9125	144
InceptionV3	0.9653	0.9000	0.9056	23
ResNet50	0.9614	0.9029	0.9076	25
Xception	0.9605	0.8977	0.9051	22

Glaucoma

ImageNet-trained CNN architectures

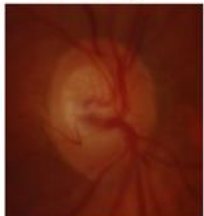

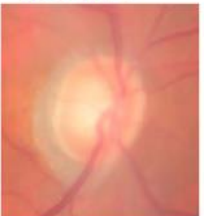


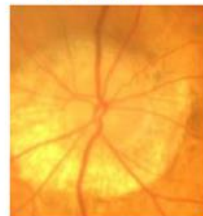
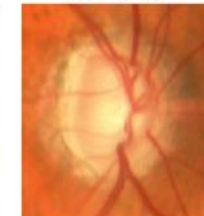
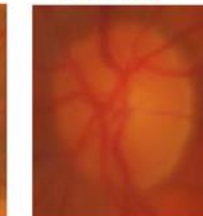





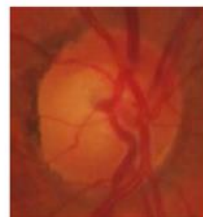
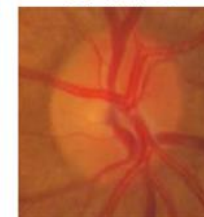
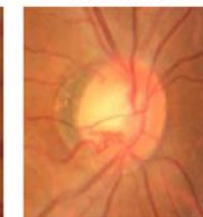
- Predicciones correctas

Ground-truth:	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
Xception score:	0.9929	0.8048	0.9837	0.9977	0.9925	0.9983	0.9905	0.9624
								
Ground-truth:	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma
Xception score:	0.9881	0.9534	0.9991	0.8964	0.9285	0.9105	0.9971	0.9975
								

Glaucoma

ImageNet-trained CNN architectures

- Predicciones incorrectas

Ground-truth:	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
Xception score:	0.0126	0.0799	0.3412	0.1306	0.1822	0.4173	0.2834	0.3487
								
Ground-truth:	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma	Glaucoma
Xception score:	0.1252	0.2853	0.1273	0.0769	0.0858	0.3459	0.1270	0.3703
								

Glaucoma

ImageNet-trained CNN architectures

- Resultados cross-testing:

Database	AUC	Accuracy	# images
HRF	0.8354	0.8000	45
Drishti-GS1	0.8041	0.7525	101
RIM-ONE	0.8575	0.7121	455
sjchoi86-HRF	0.7739	0.7082	401
ACRIMA	0.7678	0.7021	705
Chen method	0.8310	NA	650
Alghamdi	NA	0.9214	2858

Agenda

- Conceptos básicos
- Casos de estudio:
 - Tag extraction
 - Beat Event Detection
- **Tutorial con Keras:**
 - Sesión 1
 - Sesión 2
 - Ejercicio 1
 - Ejercicio 2

Tutorial con Keras

- Emplearemos Keras para aplicar transfer learning de distintos modelos a distintos conjuntos de datos.
- Pero comencemos primero jugando con el modelo InceptionV3 con TensorFlow...
- Pasos:
 - Descargar el notebook y abrirlo desde jupyter
 - Si se quiere usar GPU, subirlo a un servidor con GPU (e.g. Google Colab)