A Tutorial on Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems

Denis Parra[1], Antonio Ossa-Guerra[1], Manuel Cartagena[1], Patricio Cerda-Mardini[2], Felipe del Río[1], Isidora Palma[1], Diego Saez-Trumper[3], and Miriam Redi[3]

1. Pontificia Universidad Católica de Chile
2. MindsDB
3. Wikimedia Foundation

21st IEEE International Conference on Data Mining

ICDM 2021
IEEE International Conference on Data Mining
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Tutorial Web site

https://ialab-puc.github.io/VisualRecSys-Tutorial-ICDM2021/

A Tutorial on Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems

This page hosts the material for our work **A Tutorial on Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems**, presented at the [21st IEEE International Conference on Data Mining (IEEE ICDM 2021)].

Schedule: 16:00-18:30, Wednesday, December 8, 2021 (GMT+13, Time in Auckland, New Zealand)

Instructors

- Denis Parra, Associate Professor, PUC Chile
- Antonio Ossa-Guerra*, MSc, PUC Chile
- Manuel Cartagena, MSc, PUC Chile
- Patricio Cerda-Mardini, MSc, PUC Chile & MindsDB
- Felipe del Río, PhD Student, PUC Chile
- Isidora Palma, MSc Student, PUC Chile
- Diego Saez-Trumper, Senior Research Scientist, Wikimedia Foundation
- Miriam Redi, Senior Research Scientist, Wikimedia Foundation

Program

Duration	Overview	Presenter(s)
30 mins	**Session 1**: Introduction to Visual RecSys, datasets and feature extraction with CNNS in Python. Wikimedia Foundation and its available research resources.	Denis Parra & Diego Saez-Trumper & Miriam Redi
20 mins	**Session 2**: Pipeline for training and testing visual RecSys in Python.	Antonio Ossa-Guerra
10 mins	BREAK	-
25 mins	**Session 3**: Visual Bayesian Personalized Ranking (VBPR) and Deep Visually-aware Bayesian Personalized Ranking (DVBPR) in Pytorch [2, 3]	Patricio Cerda-Mardini
20 mins	**Session 4**: CuratorNet in Pytorch [1]	Manuel Cartagena
20 mins	**Session 5**: Attentive Collaborative Filtering (ACF) in Pytorch [4]	Felipe del Río
15 mins	Live demo of this repository	Isidora Palma
10 mins	Conclusions	Denis Parra

Tutorial Table of Contents (Starting at 16:00 NZ Time)

(30 mins) **Session 1**: (A) Intro to Visual RecSys and (B) Intro to Wikimedia resources

(20 mins) **Session 2**: Pipeline for training and testing visual RecSys in Pytorch, application with VisRank

(10 mins) [BREAK]

(25 mins) **Session 3**: Dynamic Visual Bayesian Personalized Ranking (DVBPR) in Pytorch

(20 mins) **Session 4**: CuratorNet in Pytorch

(20 mins) **Session 5**: Attentive Collaborative Filtering (ACF) in Pytorch

(15 mins) **Session 6: Demo**

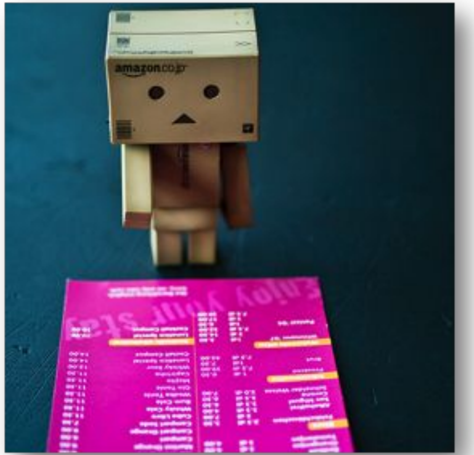
(10 mins) **Conclusion**

Session 1: Table of Contents

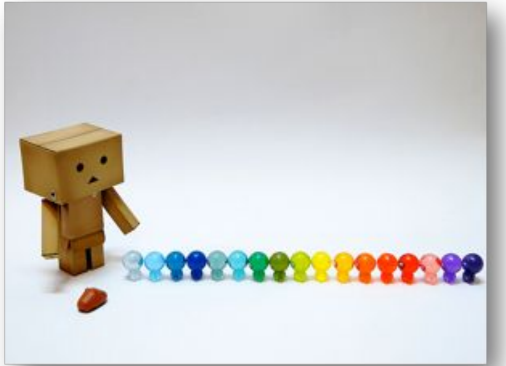
1. Introduction to visual recommender systems
2. Motivation
3. Application domains
4. Traditional approaches: Manually-engineered visual features
5. Deep Convolutional neural networks (CNNs): AlexNet, VGG, ResNet
6. Is there transfer learning from Visual Classifiers to Visual RecSys ?
7. Datasets: Is there a Movielens for visual recommendation systems?
8. The Wikimedia Commons Datase

Recommender Systems

Systems that help (groups of) people to find relevant items in a crowded item or information space (MacNee et al. 2006)



http://www.flickr.com/photos/dongga/4597533223/sizes/m/



http://www.flickr.com/photos/meaganmakes/6769496875/sizes/m/

Types of Recommender Systems

Without covering all possible methods, the two most typical classifications on recommender algorithms are:

Classification 1	Classification 2
- Collaborative Filtering - Content-based Filtering - Hybrid	- Memory-based - Model-based

Types of Recommender Systems

In this tutorial, we are focusing on these:

Classification 1	Classification 2
- Collaborative Filtering - **Content-based Filtering** - **Hybrid**	- Memory-based - **Model-based**

Motivation

Why this Tutorial on Visual Recommendation Systems?

- **Increasing growth of multimedia usage on the Web** (Images, Video)

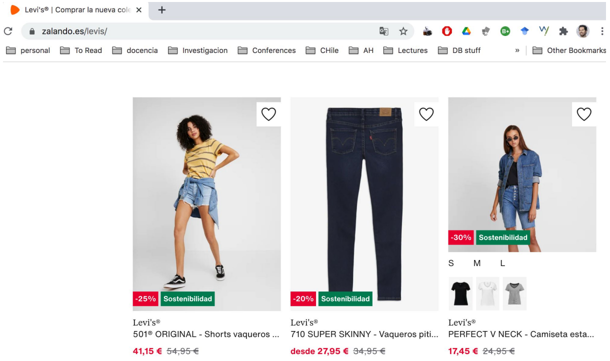
  Facebook (2004), Twitter (2006), **Pinterest** (2010), **Instagram** (2010), **TikTok** (2016)

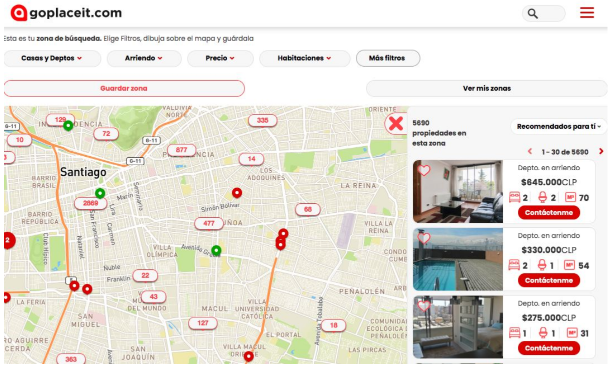
Why this Tutorial on Visual Recommendation Systems?

- Increasing growth of multimedia usage on the Web (Images, Video)

  Facebook (2004), Twitter (2006), Pinterest (2010), Instagram (2010), TikTok (2016)

- **Increasing use of multimedia for e-commerce applications** (fashion, tourism, real estate) e.g. Zalando, goplaceit





Why this Tutorial on Visual Recommendation Systems?

- Increasing growth of multimedia on the Web (Images, Video)

  Facebook (2004), Twitter (2006), Pinterest (2010), Instagram (2010), TikTok (2016)

- Increasing use of multimedia for e-commerce applications (fashion, tourism, real estate)
- **Increasing performance of visual features obtained from Deep Learning methods for related tasks (2012 - ...)**

Why this Tutorial on Visual Recommendation Systems?

- Increasing growth of multimedia on the Web (Images, Video)

  Facebook (2004), Twitter (2006), Pinterest (2010), Instagram (2010), TikTok (2016)

- Increasing use of multimedia for e-commerce applications (fashion, tourism, real estate)
- Increasing performance of visual features obtained from Deep Learning methods for related tasks (2012 - ...)

- **Growing potential of transfer learning** (2012 - ...)

Computer Vision: Historical Datasets

- 1996: faces and cars 14,000 images of 10,000 people
- 1998: MNIST 70,000 images of handwritten digits
- 2004: Caltech 101, 9,146 images of 101 categories
- 2005: PASCAL VOC 20,000 images with 20 classes

Imagenet dataset

- Imagenet [0]: Presented in 2009 at CVPR

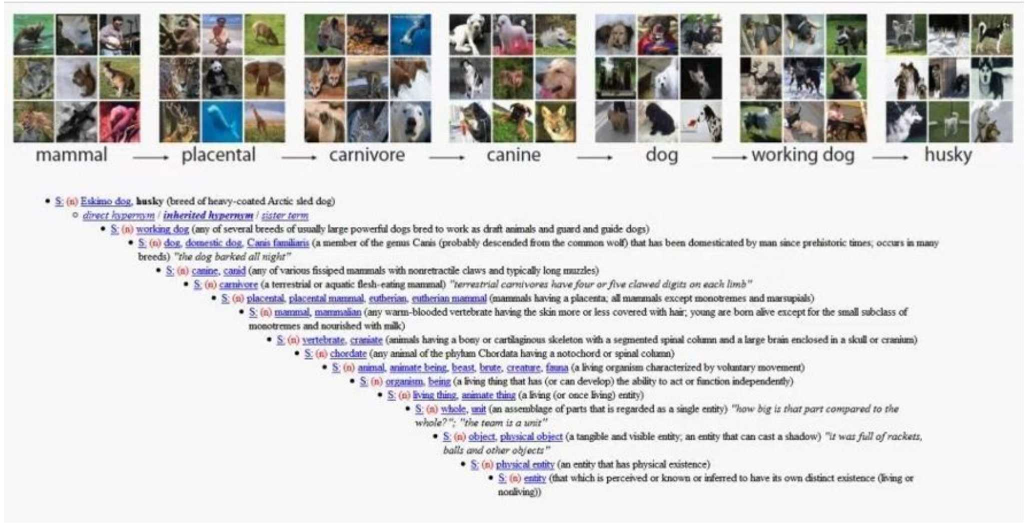
IMAGENET Large Scale Visual Recognition Challenge

- Crowdsourced
- 14,197,122 images

- 21,841 categories (non-empty synsets)
- Categories based on ***WordNet*** taxonomy

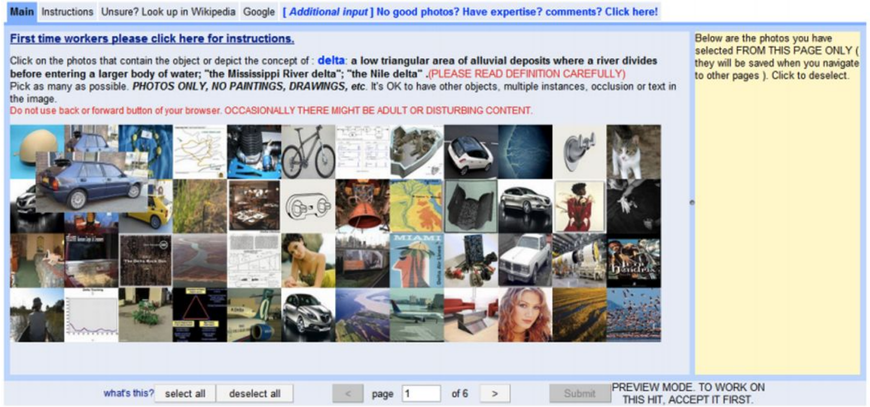
WordNet

- Wordnet: Miller’s project started in 1980 at Princeton, a hierarchy for the English language
- **Prof. Fei-Fei Li** (UIUC, Princeton, Stanford), worked on filling WordNet with many images.



Imagenet: Crowdsourced

- Amazon Mechanical Turk



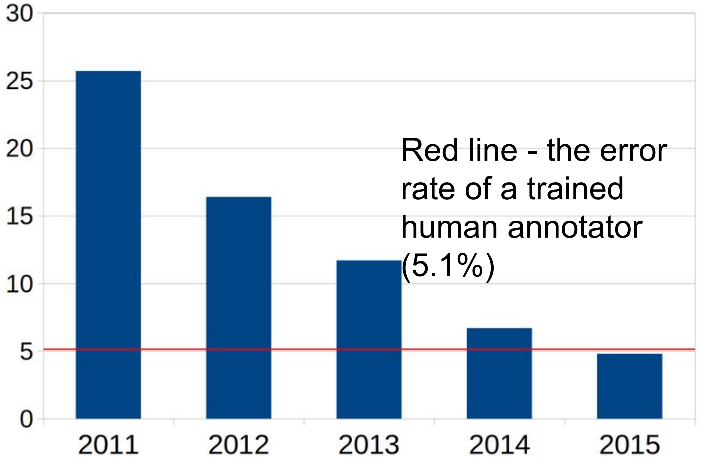
- It took 2.5 years to complete.

- Originally 3.2 million images in 5,247 categories (mammal, vehicle, etc.)

Imagenet Challenge

- The dataset was used to set a competition for image classification.

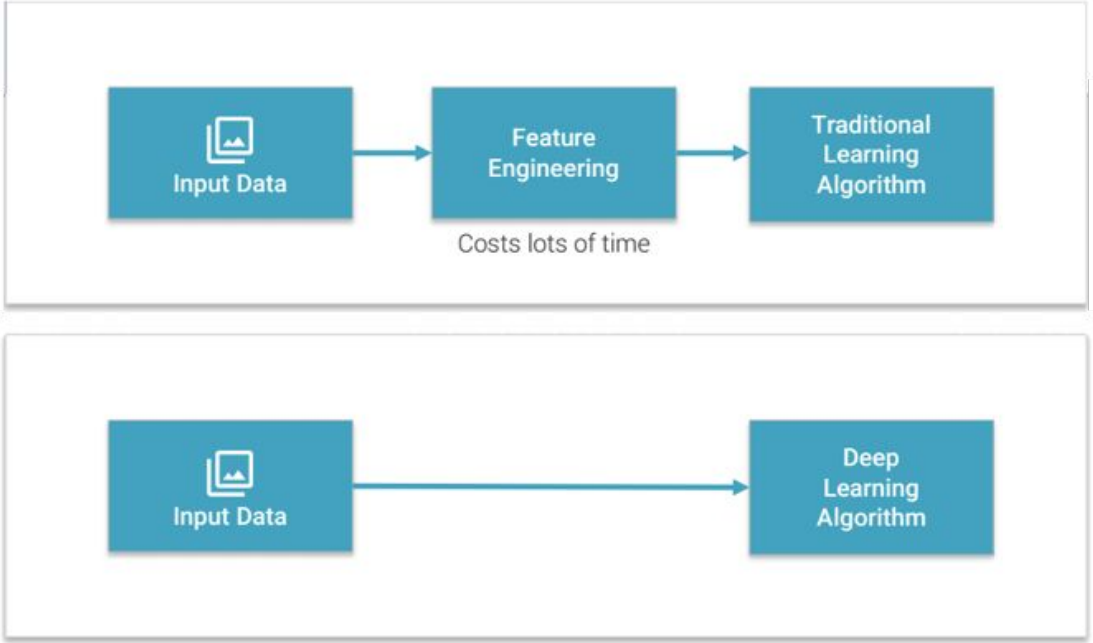
- In **2012 a team used deep learning, got error rate below 25%** (Hinton et al.), 10.8 point margin, 41% better than next best.



Progress in machine classification of images: the error rate (%) of the ImageNet competition winner by year.

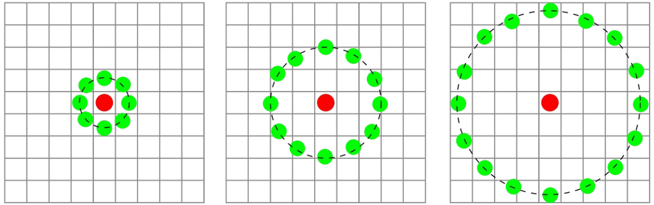
Deep Neural Networks: Representation Learning

Deep learning makes a great revolution not only in performance, but also on representation learning



Computer Vision: Historical Feature Extraction

- **[1] LBP**: Local Binary Patterns



By Xiawi - Own work, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=11743214

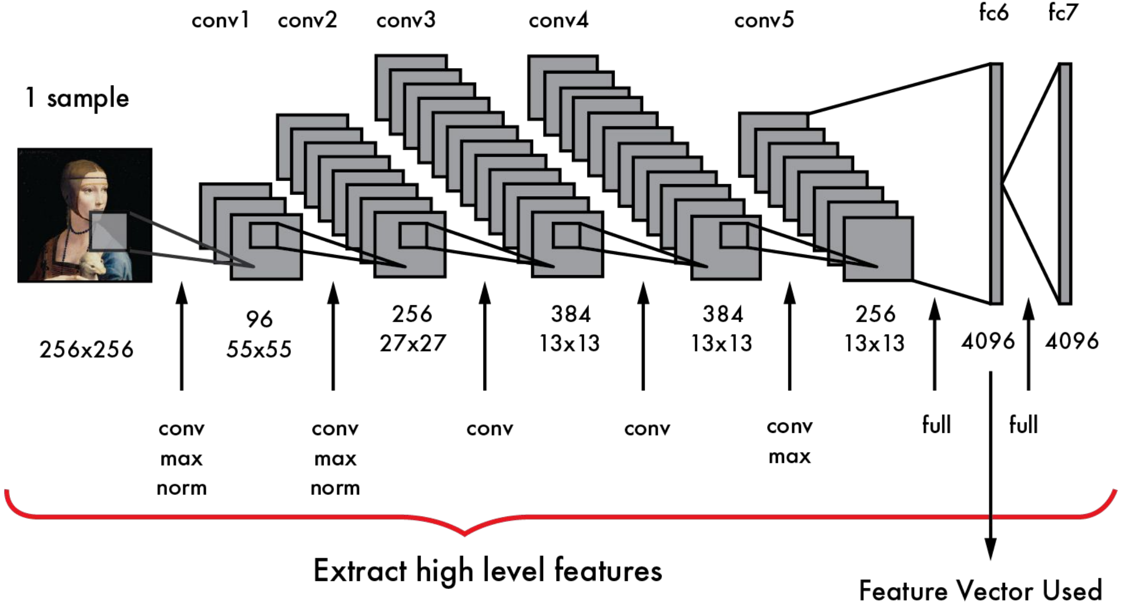
- **[2] HOG**: Histogram of Oriented Gradients

- **[3] SIFT**: scale-invariant feature transform

Deep Neural Networks: Representation Learning

AlexNet [4] made a great revolution not only in performance, but also on representation learning

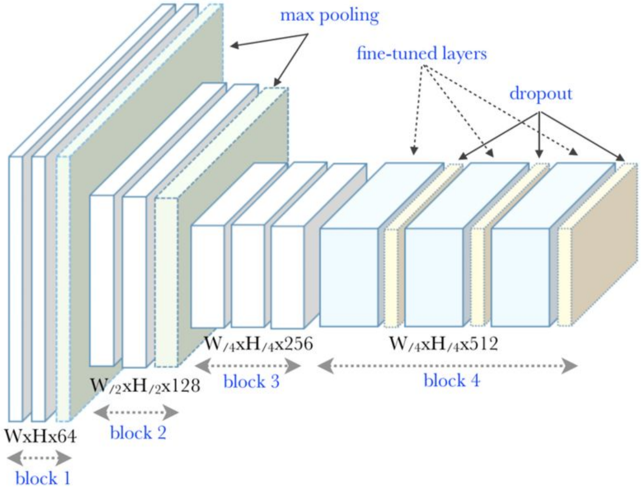
Aprox 60 million parameters



Deep Neural Networks in Computer Vision

VGG [5] introduced by the Visual Geometry Group at Oxford University, increases network depth by using very small convolution filters (3x3) compared to AlexNet. There are different versions depending on the number of layers (VGG-16/19)

Aprox 138 million parameters



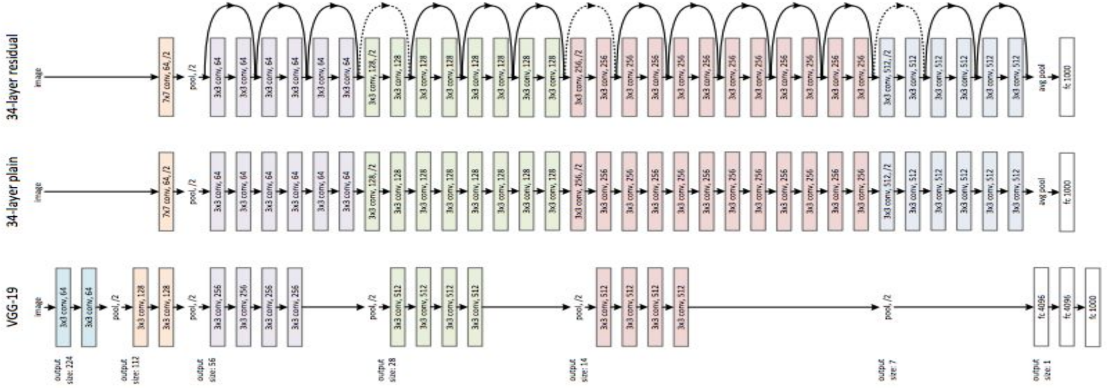
Deep Neural Networks in Computer Vision

ResNet [6] introduces a residual learning framework to ease the training of networks that are substantially deeper than those used previously (AlexNet, VGG)

ResNet-18 Aprox. 11 million parameters

method	top-1 err.	top-5 err.
VGG [40] (ILSVRC'14)	-	8.43†
GoogLeNet [43] (ILSVRC'14)	-	7.89
VGG [40] (v5)	24.4	7.1
PReLU-net [12]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

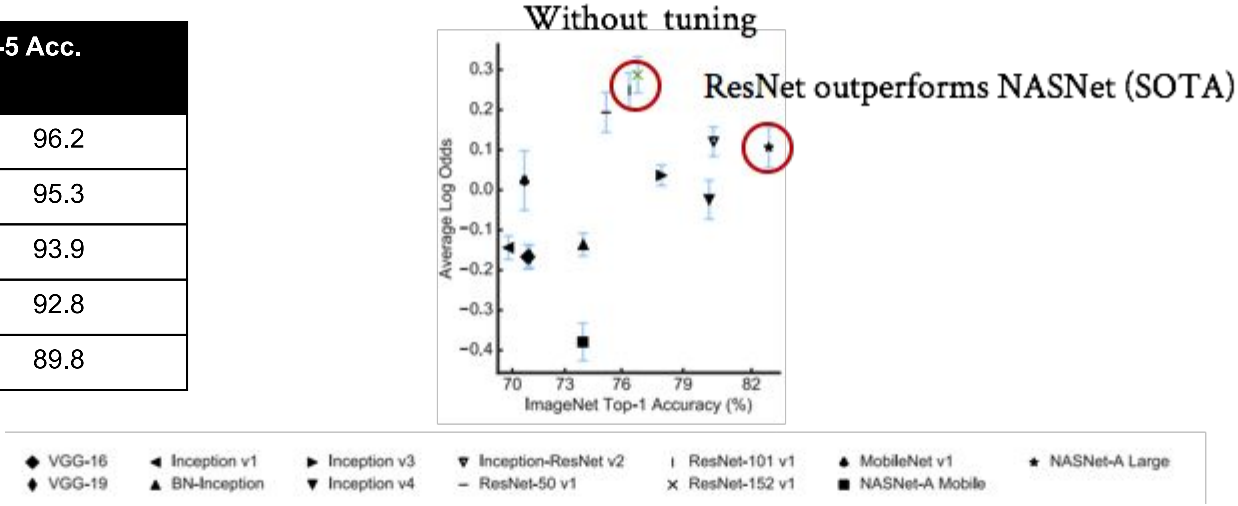


What about transfer learning?

Simon Kornblith, Jonathon Shlens, and Quoc V. Le. 2018. Do Better ImageNetModels Transfer Better? (2018). https://arxiv.org/abs/1805.08974

Method	Top-1 Acc	Top-5 Acc.
NASNet Large	82.7	96.2
InceptionResNetV2	80.4	95.3
InceptionV3	78.0	93.9
ResNet50	75.6	92.8
VGG19	71.1	89.8

https://github.com/tensorflow/models/tree/master/research/slim#pre-trained-models



What about transfer learning for Visual RecSys?

Using pre-trained neural networks, there is not correlation between Imagenet and image recsys performance [7].

CNN	Artwork Image Recommendation				ILSVRC-2012-CLS	
	R@20	P@20	MRR@20	nDCG@20	Top-1 Acc. (%)	Top-5 Acc. (%)
ResNet50	**.1632**	**.0141**	**.0979**	**.1253**	75.2	92.2
VGG19	**.1398**	**.0124**	**.0750**	**.1008**	71.1	89.8
NASNet Large	.1379	.0120	.0743	.0998	**82.7**	**96.2**
InceptionV3	.1332	.0125	.0744	.1007	78.0	93.9
InceptionResNetV2	.1302	.0117	.0692	.0936	**80.4**	**95.3**
Random	.0172	.0013	.0051	.0093	-	-

Datasets for Visual Recommender Systems

Is there a **Movielens** dataset to train and benchmark visual recommendation systems ?

Datasets for Visual Recommender Systems

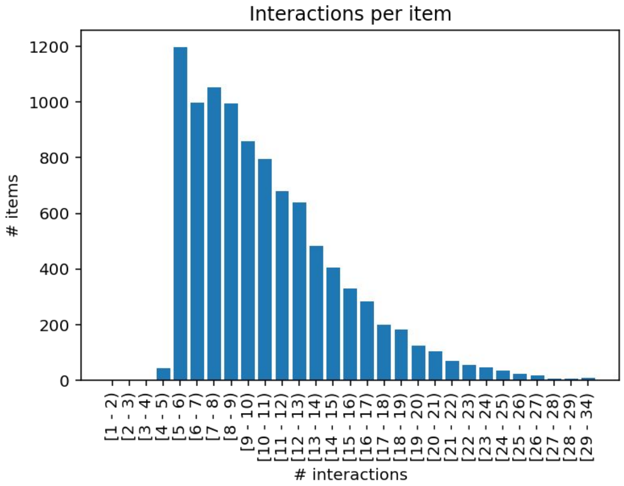
Is there a **Movielens** dataset to train and benchmark visual recommendation systems ?

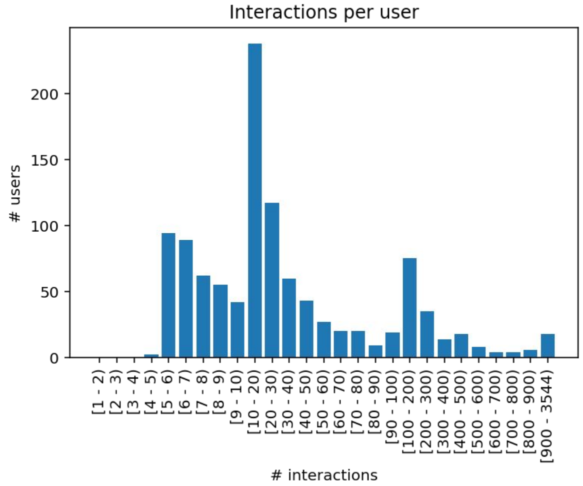
Not exactly. There are some datasets, but usually you find embeddings (npy files) but not images, or the URL to files you need to download on your own

- https://cseweb.ucsd.edu/~jmcauley/datasets.htm (Behance, Amazon)
- Pinterest, mongoDB dataset (https://goo.gl/LjMoYa)
- UGallery (provided by us at https://github.com/ialab-puc/CuratorNet)

The Wikimedia Commons Dataset

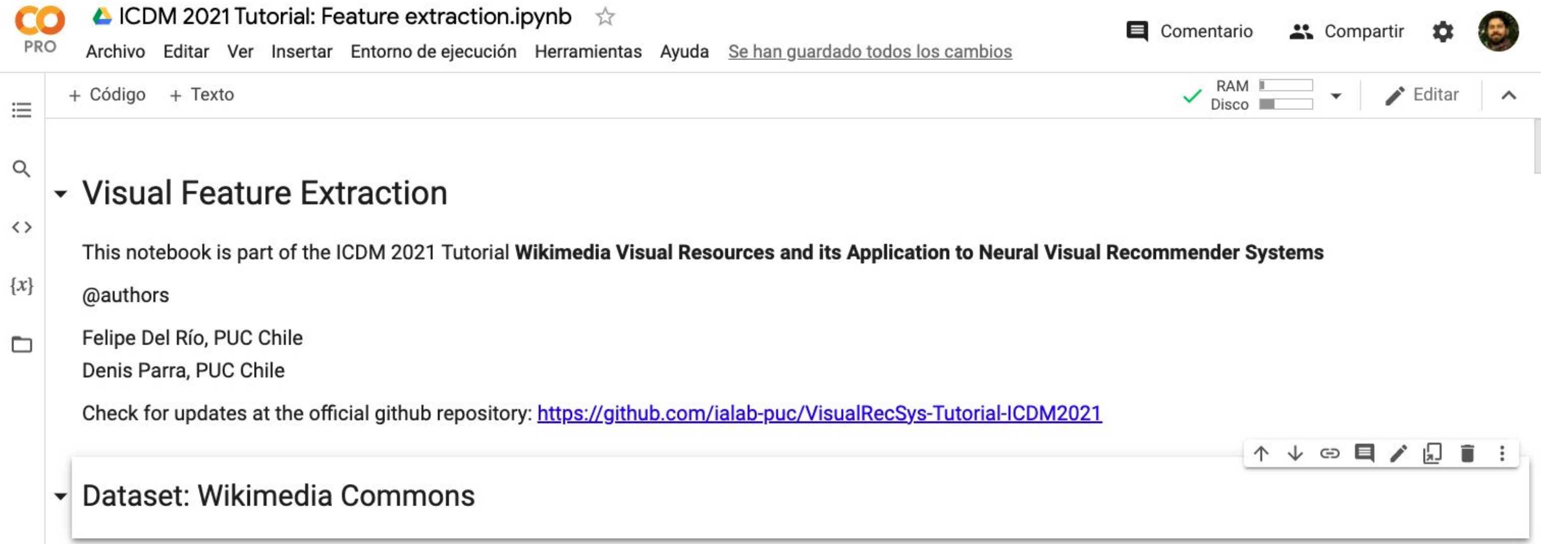
- Thanks to Miriam Redi and Diego Saez from Wikimedia Foundation
- We share a sample for the community
  - 1,079 unique users / 9,636 (32,958) unique items / 96,991 interactions





Visual Feature Extraction from a pre-trained CNN

https://colab.research.google.com/drive/1JCTPS88AzKA0KNVCoEvYCBaaYebgdoYn



References

[0] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[1] He, D. C., & Wang, L. (1990). Texture unit, texture spectrum, and texture analysis. IEEE transactions on Geoscience and Remote Sensing, 28(4), 509-512.

[2] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893).

[3] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157).

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[5] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[7] del Rio, F., Messina, P., Dominguez, V., & Parra, D. (2018). Do Better ImageNet Models Transfer Better... for Image Recommendation?. arXiv preprint arXiv:1807.09870.

Thank you!

Denis Parra [dparra@ing.puc.cl]

Denis Parra[1], Antonio Ossa-Guerra[1], Manuel Cartagena[1], Patricio Cerda-Mardini[2], Felipe del Río[1], Isidora Palma[1], Diego Saez-Trumper[3], and Miriam Redi[3]

1. Pontificia Universidad Católica de Chile
2. MindsDB
3. Wikimedia Foundation