lecture 4: shooting
part 2

daniel gatica-perez
announcements

reading #5 will be presented today

A. G. Reece and C. M. Danforth, Instagram photos reveal predictive markers of depression, EPJ Data Science, 2017
reminder: project schedule

1. team building: DONE
   email the list of your team members on Fri 05.03.2021
each team will have a designated project mentor

2. project pitch: DONE
   5-minute presentation of your project on Fri 26.03.2021
   structure: title, problem, goals, approach, evaluation

3. project progress presentation on Fri 30.04.2021
   5-minute presentation per team about project progress

4. final project presentation on Fri 11.06.2021
   talk: 25-minute presentation + 20-minute questions
   program: 09:00-16:30

5. final project report by Fri 18.06.2021
   ACM conference paper format (6 pages + references + appendix)
final project presentation day (friday 11.06.2021)

09:00-09:45   group 1
09:45-10:30   group 2
10:30-10:45   break
10:45-11:30   group 3
11:30-12:15   group 4

12:15-13:00  lunch break

13:00-13:45   group 5
13:45-14:30   group 6
14:30-14:45   break
14:45-15:30   group 7
15:30-16:15   group 8

+ everybody is encouraged to attend the full day
+ please reserve the slot for your team
this lecture

1. a snapshot of the present
   - flickr, instagram, snapchat
2. a look into the past
   - 20th century image sharing practices
3. understanding research on social image systems
   - **flickr**: tags & communities
     - computer vision as an enabler
   - **instagram**: selfies & engagement
   - **snapchat**: ephemeral social media
tags, communities, geo-tagged media

selfies & followers

ephemeral social media
computer vision as an enabler
A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

A herd of elephants walking across a dry grass field.

A close up of a cat laying on a couch.

Show and tell: A neural image caption generator, O Vinyals, A Toshev, S Bengio, D Erhan, CVPR 2015
visual classification with deep learning imagenet data


Figure 3. Example network architectures for ImageNet. **Left:** the VGG-19 model [40] (19.6 billion FLOPs) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

what should computers recognize?
who decides?
before: corel images & object labels

Corel Gallery (1st ed. 1994)

https://openlibrary.org/works/OL16047006W/Corel_gallery

airplanes, black bears, brown bears, cheetahs, eagles, elephants, horses, polar bears, tigers, zebras.

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.
today: 365 scene categories

By MIT Computer Science and Artificial Intelligence Laboratory

Scene recognition is one of the hallmark tasks of computer vision, allowing defining a context for object recognition. Here we introduce a new scene-centric database called Places, with 205 scene categories and 2.5 millions of images with a category label. Using convolutional neural network (CNN), we learn deep scene features for scene recognition tasks, and establish new state-of-the-art performances on scene-centric benchmarks. Here we provide the Places Database and the trained CNNs for academic research and education purposes.

Announcement

- **NEW** Places2, the 2nd generation of the Places Database, is available for use, with more images and scene categories. CNNs trained on Places365 (new Places2 data) are also released.
- Scene Parsing Challenge 2016 and Places Challenge 2016 are hosted at ECCV'16.
- Places205-VGG and Places205-GoogLeNet are available to download in the Places CNNs.

http://places.csail.mit.edu/
before: human labels

Figure 7: Output of our face detector on a number of test images from the MIT+CMU test set.

P Viola, M Jones, Rapid object detection using a boosted cascade of simple features, in Proc. CVPR 2001
today: biometrics as a service

Face detection and analysis

With Amazon Rekognition, you can easily detect when faces appear in images and videos and get attributes such as gender, age range, eyes open, glasses, facial hair for each. In video, you can also measure how these face attributes change over time, such as constructing a timeline of the emotions expressed by an actor. Learn more »

Face search and verification

Amazon Rekognition provides fast and accurate face search, allowing you to identify a person in a photo or video using your private repository of face images. You can also verify identity by analyzing a face image against images you have stored for comparison. Learn more »

https://aws.amazon.com/rekognition/
who labels the images?
Welcome to LabelMe, the open annotation tool.

The goal of LabelMe is to provide an online annotation tool to build image databases for computer vision research. You can contribute to the database by visiting the annotation tool.
Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it’s convenient.

249,109 HITs available. View them now.

Make Money by working on HITs

HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:
- Can work from home
- Choose your own work hours
- Get paid for doing good work

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you’re satisfied with the results

Find more about being a Worker
what was Instagram at the beginning?

mobile app with image filters
vintage-looking images

credit: instagram.com/press
what do people share on Instagram?
selfie (n.)

“a photograph that one has taken of oneself, typically one taken with a smartphone or webcam and uploaded to a social media website”
“SELFIE” BEFORE SMART PHONES

Van Gogh  Tintoretto  Rembrandt  Tiziano

Rubens  Renoir  Gauguin  Monet

Da Vinci  Kahlo  Degas  Rivera

https://twitter.com/vangoghmuseum/status/452884195208359936/photo/1

Don’t miss what’s happening
what’s shared on Instagram?

1000 photos = 50 **active** users ( >30 followees, >30 followers, >60 photos); 20 photos per user

Random 200-photo sample:
+ extract visual features (SIFT)
+ get 15 clusters (k-means)
+ manually refine clusters
+ 8 final categories

Remaining 800 photos
+ map to closest category
+ single category per photo

<table>
<thead>
<tr>
<th>Category</th>
<th>Exemplary Photos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friends (users posing with others; At least two human faces are in the photo)</td>
<td><img src="image" alt="Friends" /></td>
</tr>
<tr>
<td>Food (food, recipes, cakes, drinks, etc.)</td>
<td><img src="image" alt="Food" /></td>
</tr>
<tr>
<td>Gadget (electronic goods, tools, motorbikes, cars, etc.)</td>
<td><img src="image" alt="Gadget" /></td>
</tr>
<tr>
<td>Captioned Photo (pictures with embed text, memes, and so on)</td>
<td><img src="image" alt="Captioned Photo" /></td>
</tr>
<tr>
<td>Pet (animals like cats and dogs which are the main objects in the picture)</td>
<td><img src="image" alt="Pet" /></td>
</tr>
<tr>
<td>Activity (both outdoor &amp; indoor activities, places where activities happen, e.g., concert, landmarks)</td>
<td><img src="image" alt="Activity" /></td>
</tr>
<tr>
<td>Selfie (self-portraits; only one human face is present in the photo)</td>
<td><img src="image" alt="Selfie" /></td>
</tr>
<tr>
<td>Fashion (shoes, costumes, makeup, personal belongings, etc.)</td>
<td><img src="image" alt="Fashion" /></td>
</tr>
</tbody>
</table>

Table 1: **8 Photo Categories**

Y. Hu, L. Manikonda, S. Kambhampati, What We Instagram: A First Analysis of Instagram Photo Content and User Types, in Proc. AAAI ICWSM 2014
Figure 2: Proportion of Categories
large-scale selfie analysis
Figure 3: Example pictures of Selfie, Alt, and Face datasets as well as features predicted by Face++.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
<th># Pictures</th>
<th># Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selfie</td>
<td>Pictures with hashtags containing ‘selfie’ (e.g., #selfie, #selfietoday)</td>
<td>1,196,080</td>
<td>214,656</td>
</tr>
<tr>
<td>Alt</td>
<td>Pictures with alternative hashtags for ‘selfie’ (e.g., #selca, #selstagram)</td>
<td>2,453,749</td>
<td>242,650</td>
</tr>
<tr>
<td>Face</td>
<td>Pictures with face(s) detected using the Face++ tool</td>
<td>1,921,207</td>
<td>315,751</td>
</tr>
<tr>
<td>All</td>
<td>Randomly chosen set of pictures</td>
<td>10,000,019</td>
<td>184,615</td>
</tr>
</tbody>
</table>

Table 1: Number of media and users in each of the four datasets used in this paper.

Figure 4: Longitudinal trend of selfie posts across the datasets.
faces get more likes on instagram
two questions:
- do photos with & without faces differ w.r.t. engagement (likes & comments)?
- do features of the subject (gender, age) affect engagement?

Accuracy (w.r.t. majority vote):
- has face: 97%
- has female face: 96%
- has male face: 96%
- has face < 18 years old: 93%
- has face in (18,35) years old: 96%
- has face > 35 years old: 99%
<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Description</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>likes*</td>
<td>Number of likes on each photo.</td>
<td><img src="image" alt="Distribution" /></td>
</tr>
<tr>
<td>(dependent)</td>
<td>comments*</td>
<td>Number of comments on each photo.</td>
<td><img src="image" alt="Distribution" /></td>
</tr>
<tr>
<td>Audience &amp; Activity</td>
<td>followers*</td>
<td>Number of users who follow the photo’s owner.</td>
<td><img src="image" alt="Distribution" /></td>
</tr>
<tr>
<td>(features)</td>
<td>photos*</td>
<td>Number of photos shared by photo’s owner.</td>
<td><img src="image" alt="Distribution" /></td>
</tr>
</tbody>
</table>

Table 1. Distributions of quantitative and binary variables used in this paper. Variables marked with **are log-transformed. The red and blue lines identify mean and median of the distribution, respectively. Orange refers to 1’s in the bar graphs. The engagement variables are our dependent measures. Audience and activity variables are used as controls, and faces variables are the focal point of this study.
Table 1. Distributions of quantitative and binary variables used in this paper. Variables marked with "*" are log-transformed. The red and blue lines identify mean and median of the distribution, respectively. Orange refers to 1’s in the bar graphs. The engagement variables are our dependent measures. Audience and activity variables are used as controls, and faces variables are the focal point of this study.
data analysis model

- Face binary features (6)
- # followers
- # photos
- # likes
- # comments

After applying log transformation, the data is normalized with z-score normalization.

The negative binomial regression parameters (beta, p-values) are used.

Model when dependent variable is count data and over-dispersed (observed variance is higher than the variance of theoretical model).
"the higher the number of followers, the more likely it is for a photo of that user to receive likes and comments" [larger audience for a given photo]

"the higher activity (number of photos), the lower chances of receiving likes and comments" [less “exposure” time for any given photo]

"the number of likes and comments are significantly higher when there is at least one face in the image” [photos with faces engage people more]
visualizing instagram: selfiecity

Investigating the style of *self-portraits* (*selfies*) in five cities across the world.

Selfiecity investigates *selfies* using a mix of theoretic, artistic and quantitative methods:

- We present our findings about the demographics of people taking selfies, their poses and expressions.
- Rich media visualizations (*imageplots*) assemble thousands of photos to reveal interesting patterns.
- The interactive *selfiexploratory* allows you to navigate the whole set of 3200 photos.
- Finally, theoretical essays discuss selfies in the history of photography, the functions of images in social media, and methods and dataset.

selfiecity.net (L. Manovich et al., 2014)
inferring real-world events from Instagram data
Feel #happy in #Gruyere. Have lunch with #cheese, #rosti at #fancy restaurant with #friends.
+ ML + CV + NLP to understand:

Obesity rates [1]
Food perception [2,3]
Food deserts [4]
Depression [5]
Mental health [6]
etc. etc.

instagram images & hashtags for deep learning
hashtags as weak labels: noisy, but abundant and for free
basic ideas

“Training image recognition networks on large sets of images with hashtags, the biggest of which included 3.5B images and 17,000 hashtags, as labels instead of manually categorizing each picture.”

“This offers important insight into how to shift from supervised to weakly supervised training, where we use existing labels — hashtags — rather than ones that are chosen and applied specifically for AI training.”

“We created a way to distribute the task across up to 336 GPUs, shortening the total training time to just a few weeks.”

https://engineering.fb.com/2018/05/02/ml-applications/advancing-state-of-the-art-image-recognition-with-deep-learning-on-hashtags/

Photo by Elena Koycheva on Unsplash
https://unsplash.com/photos/yJwbvWmJs5M
### 1B+ Instagram Images for Weakly Supervised Pretraining

<table>
<thead>
<tr>
<th>Name template</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>train-IG-I-1.5k</td>
<td>Instagram training set of $I$ images and $\sim1.5k$ hashtags from ImageNet-1k.</td>
</tr>
<tr>
<td>train-IG-I-8.5k</td>
<td>Instagram training set of $I$ images and $\sim8.5k$ hashtags from WordNet.</td>
</tr>
<tr>
<td>train-IG-I-17k</td>
<td>Instagram training set of $I$ images and $\sim17k$ hashtags from WordNet.</td>
</tr>
<tr>
<td>train-IN-1M-1k</td>
<td>The standard ImageNet-1k ILSVRC training set with 1.28M images.</td>
</tr>
<tr>
<td>val-IN-50k-1k</td>
<td>The standard ImageNet-1k ILSVRC validation set with 50k images.</td>
</tr>
<tr>
<td>train-IN-I-L</td>
<td>Extended ImageNet training set of $I$ images and $L \in {5k, 9k}$ labels.</td>
</tr>
<tr>
<td>val-IN-I-L</td>
<td>Extended ImageNet validation set of $I$ images and $L \in {5k, 9k}$ labels.</td>
</tr>
<tr>
<td>train-Places-1.8M-365</td>
<td>The Places365-Standard training set (high-resolution version).</td>
</tr>
<tr>
<td>train-COCO-115k-80</td>
<td>The standard COCO detection training set (2017 version).</td>
</tr>
<tr>
<td>val-COCO-5k-80</td>
<td>The standard COCO detection validation set (2017 version).</td>
</tr>
<tr>
<td>test-COCO-20k-80</td>
<td>The standard COCO detection test-dev set (2017 version).</td>
</tr>
</tbody>
</table>

**Table 1: Summary of Image Classification Datasets.** Each dataset is named with a template, `role-source-I-L`, that indicates its role (training, validation, testing), source, number of images $I$, and number of labels $L$.

D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, L. van der Maaten, Exploring the Limits of Weakly Supervised Pretraining, in Proc ECCV 2018.
results

Fig. 1: Classification accuracy of ResNeXt-101 32×16d pre trained on IG-1B with different hashtag vocabularies (purple bars) on IN-{1k, 5k, 9k} (left) and CUB2011, Places365 (right). Baseline models (gray bars) are trained on IN-{1k, 5k, 9k} (left) and IN-1k (right), respectively. Full network finetuning is used. Higher is better.
SEER: The start of a more powerful, flexible, and accessible era for computer vision

“After pretraining on a billion random, unlabeled and uncurated Instagram images, SEER outperformed the state-of-the-art self-supervised systems, reaching 84.2 percent top-1 accuracy on ImageNet.”

March 4, 2021
tags, communities, geo-tagged media

selfies & followers

ephemeral social media
ephemeral image sharing
affordances

“possible actions a person can perform on an object”
(Don Norman)

credit: http://www.flickr.com/photos/pocait/2634190989 (cc)
four affordances of social media

“persistence:
the durability of online expressions & content

visibility:
the potential audience who can bear witness

spreadability:
the ease with which content can be shared

searchability:
the ability to find content”

persistence

it enables asynchronous interaction

conversations endure; messages don’t expire

posts are "on the record" forever

the opposite of ephemeral

“When a user wants to send a picture to a friend (a **snap**), they use their phone camera to take a picture from inside the app.”

“Senders choose a receiver, and can customize the snap by adding a brief caption or drawing on it.”

“They can also set the lifespan of a snap, how long the receiver has after opening it before the picture is automatically deleted (1-10 sec.)”

**Story** function allows users to send snaps to their whole network. Story snaps last for 24 hours, when any friend of the user can view the snap.”

---

“Ephemerality strongly emphasizes the affordance of default deletion, while screenshots afford selective saving with notification. These affordances are different than other media, where persistence typically affords recordability, reviewability, and replicability“

“The disappearance of images may (or may not) afford a new kind of privacy

What makes Snapchat matter has to do with how it treats attention”
"When someone sends an image via Snapchat, they choose how long you get to view the image/video. The underlying message is simple: You’ve got a few seconds. PAY ATTENTION.

 Teens choose not to open a Snap the moment they get it because they want to wait for the moment when they can appreciate it… And when they do, they just concentrate on what’s in front of them.

 Snapchat invites focus … it is a reminder that … the ephemeral is valuable.”

[Instagram Stories copied the model (image disappearance after 24 hours) in Aug. 2016]
“Smaller, More Intimate Networks”
“E-mails are for professors […]. Text messages are for my family or my best friend who I can always reach out to […]. Snapchat is definitely for just my age group, especially ones who are close to me and who know me very well. (P16)”

“Everyday Talk with Close Relationships”
“[Snapchat interactions mostly are] just one or two snaps back and forth, you see their face, you exchange a laugh even though it’s not like personally … A little with just keeping connected but it’s I think it’s kind of on superficial level. (P11)”

“Deletion Makes Space for the Mundane”
“I won’t look back at someone’s old photos. I don’t do that frequently. I’m just interested in the moment and I don’t care about it after I see it. (P23)”

“Performance with Less Self-consciousness”
“There are definitely things on Snapchat that people will video or take a picture of me that I wouldn’t want on Instagram or Facebook. (P10)”

“Ephemerality Mitigates Long-term Exhibition Concerns”
“You need to be much more careful about how you use Facebook messenger than Snapchat, because [on] Facebook stories, you can scroll through the log with everyone you ever talked to on Facebook, and look what was said. (P7)”

what to remember

Social images as community activity
Tags and communities as key features

Social images as expression of the self
Computer vision enable detailed analyses
Higher attention when faces are present

Social images as ephemeral activity
Attention is focused through ephemerality
Default deletion & selective saving enable other practices
questions?

daniel.gatica-perez@epfl.ch