GREC2017 FINAL PANEL DISCUSSION

12th IAPR International Workshop on Graphics Recognition

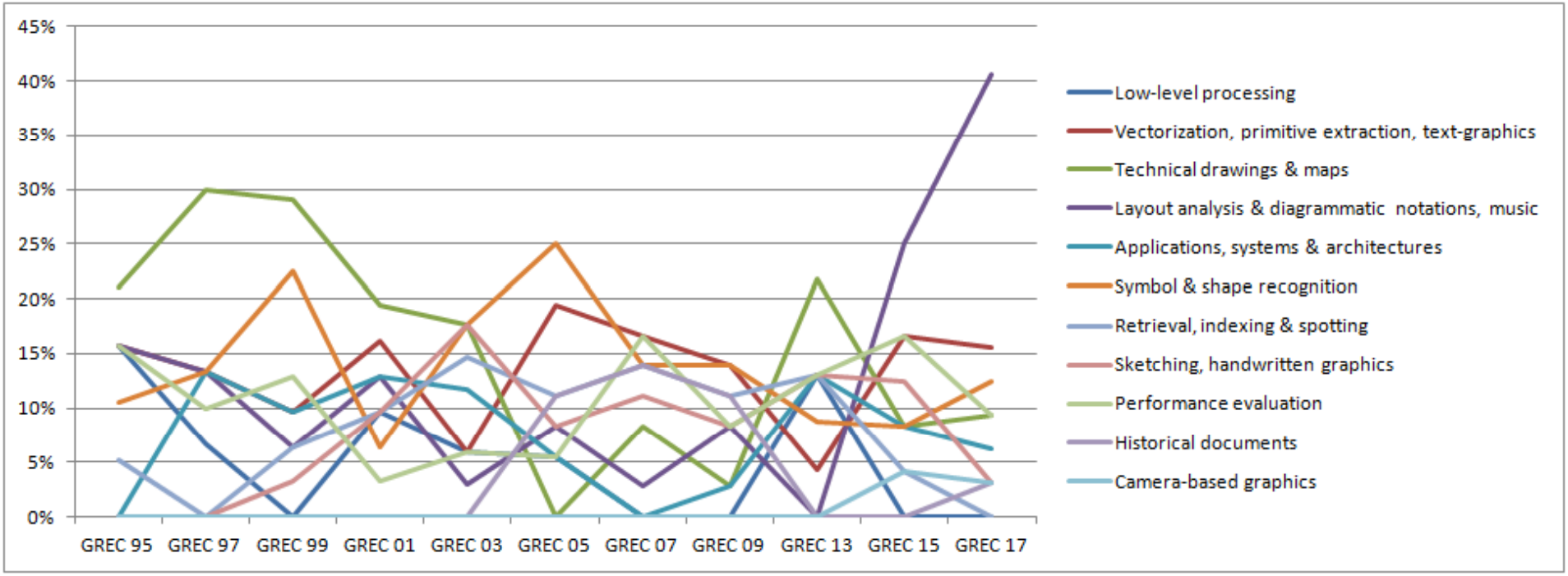
Kyoto, Japan - November 9-10 2017

Josep Lladós

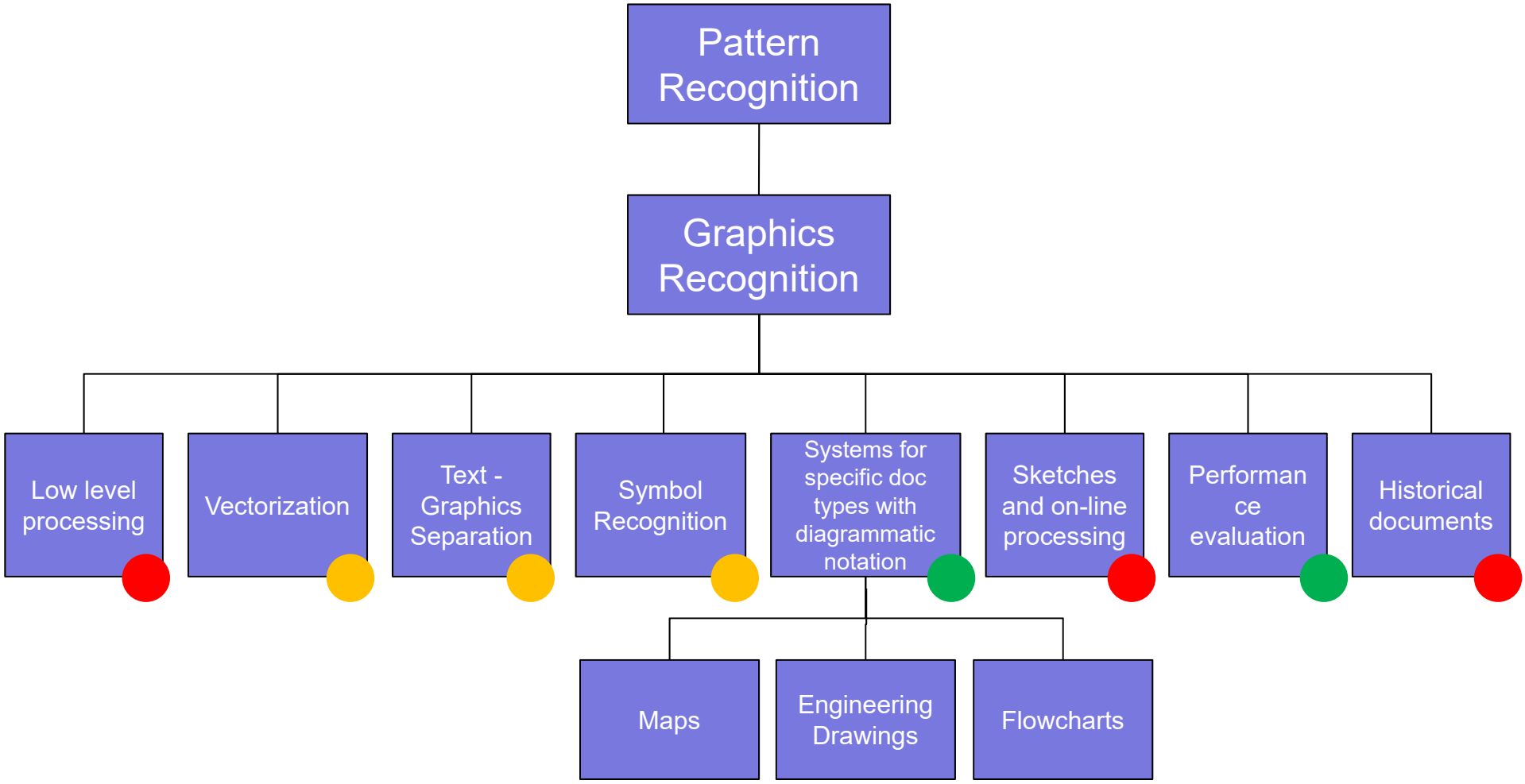
Statistics in GREC series

	GREC 95	GREC 97	GREC 99	GREC 01	GREC 03	GREC 05	GREC 07	GREC 09	GREC 13	GREC 15	GREC 17
Low-level processing	16%	7%	0%	10%	6%	6%	0%	0%	13%	0%	0%
Vectorization, primitive extraction, text-graphics	16%	13%	10%	16%	6%	19%	17%	14%	4%	17%	16%
Technical drawings & maps	21%	30%	29%	19%	18%	0%	8%	3%	22%	8%	9%
Layout analysis & diagrammatic notations, music	16%	13%	6%	13%	3%	8%	3%	8%	0%	25%	41%
Applications, systems & architectures	0%	13%	10%	13%	12%	6%	0%	3%	13%	8%	6%
Symbol & shape recognition	11%	13%	23%	6%	18%	25%	14%	14%	9%	8%	13%
Retrieval, indexing & spotting	5%	0%	6%	10%	15%	11%	14%	11%	13%	4%	0%
Sketching, handwritten graphics	0%	0%	3%	10%	18%	8%	11%	8%	13%	13%	3%
Performance evaluation	16%	10%	13%	3%	6%	6%	17%	8%	13%	17%	9%
Historical documents	0%	0%	0%	0%	0%	11%	14%	11%	0%	0%	3%
Camera-based graphics	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	3%

Statistics in GREC series

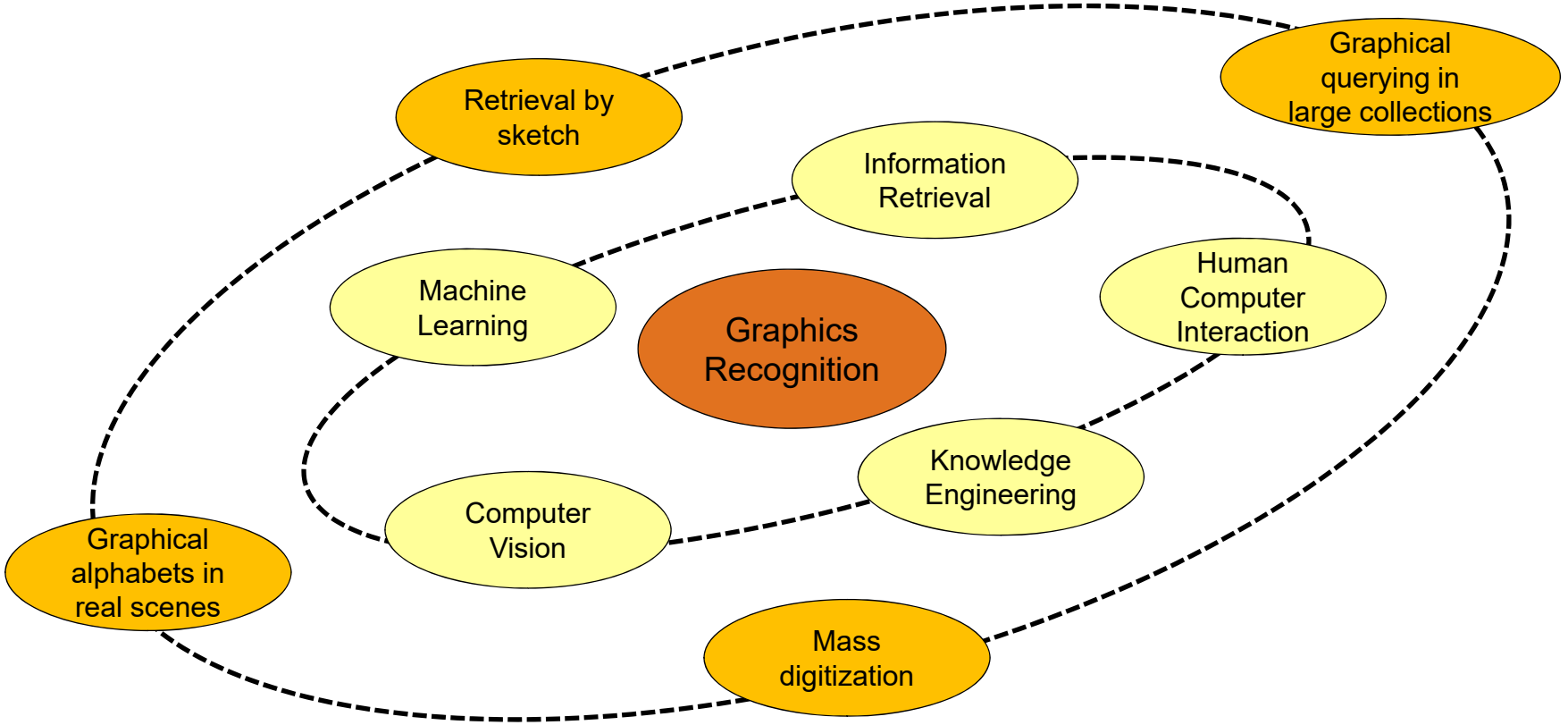


A traditional view of Graphics Recognition



Graphics Recognition: an egocentric perspective

- An orbital view: GR is a croosroad, it eventually attracts researchers from many areas to solve problems.
- The “glue” is just the need to understand graphical pieces of information that were made by humans to be read by humans in a number of applications and society needs.



In GREC2017 we noticed that

- The traditional steps (vectorization, text/graphics separation, symbol recognition) are still there but ...
  - They are losing strength by themselves, but they make sense in a global pipeline.
  - If we analyze them individually, the state of the art is close to consider the problems are solved.
- The inclusion of traditional topics in a broader context (e.g. music scores, diagrams, engineering drawings, maps) is more challenging.

Conclusion 1? GR is a component in end-to-end interpretation systems (machines as “message decoders” where graphical languages are an important but not unique component).

GR in more global end-to-end systems

- Need to incorporate more semantics: language and context.
- Need to cope with genericity and heterogeneity: GR as a service that should be “offered” to several interpretation pipelines.
- Need to make systems scalable: large scale interpretation.

Conclusion 2? As researchers, there is a need to escape from our confort zone, where we are designing ad-hoc methods for particular problems.
From a semiotic point of view, the field will move from the signifier (recognition of the compounding symbols) to the significant, i.e. the reading and understanding of the sign system in the context it appears.

GR in the Deep Era

- As in the other areas, Deep Neural Networks have irrupted in GR.
- But is it the silver bullet? Do we really need it for everything? Take into account the cost of learning (data).
- Graphical documents involve 2D visual languages.

Conclusion 3? As in textual objets (OCR, HTR, NLP) language models have been integrated in deep learning architectures, the integration of bidimensional language models is a challenge for the next years.

The need for anotated data

- Big amount of ground truth data is required, not only for performance evaluation, but also for training.
- In addition to classical ways of generating data (crowdsourcing) there are new challenging directions: data augmentation, synthetic generation.

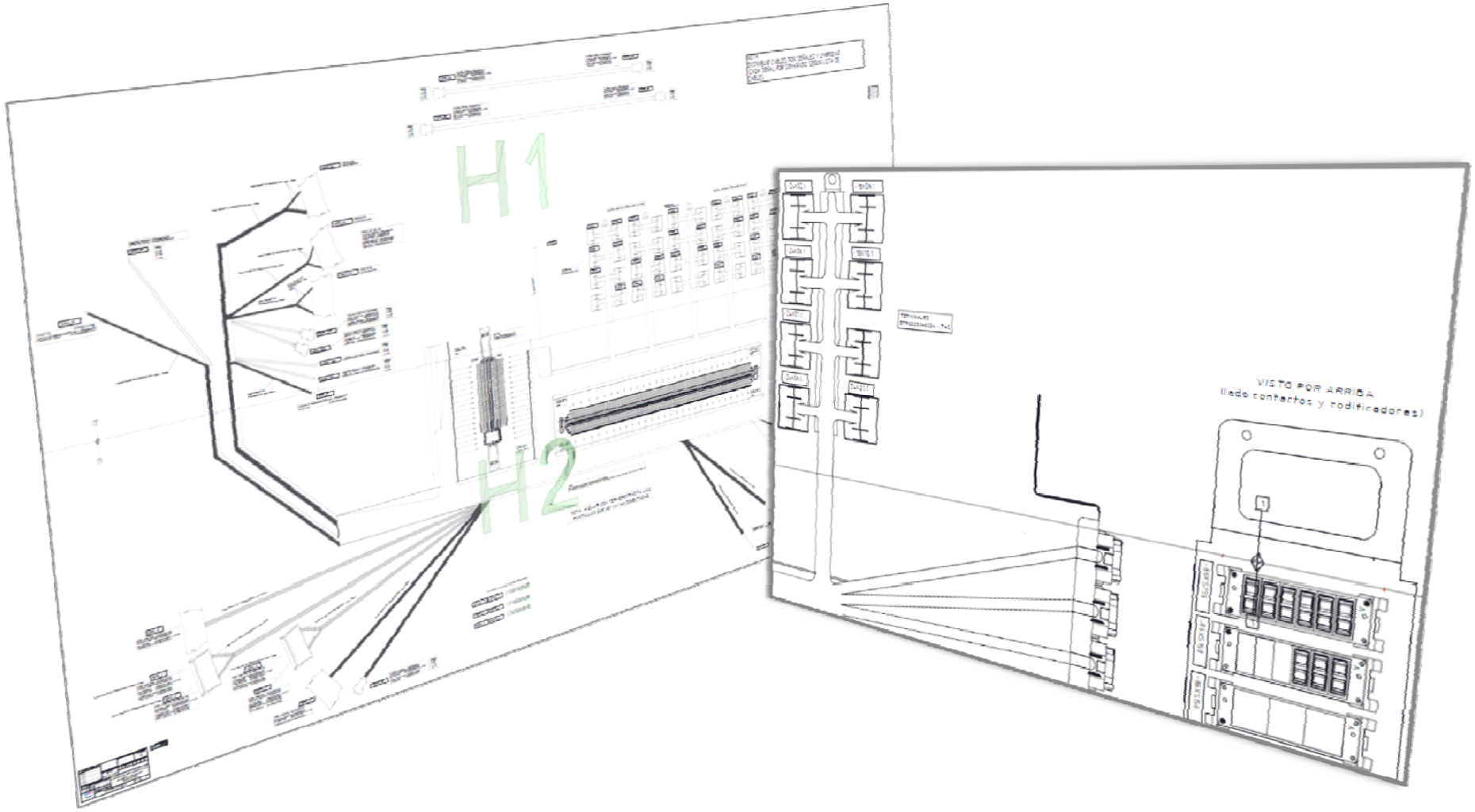
Conclusion 4? We have to take advantage of the effort made by the community and centralise data and protocols (e.g. the Engineering Drawings Challenge) → The role of the TC10/TC11 dataset curators in defining the roadmap for data generation.

What is being done that involves GR?

- In addition to the traditional topics that we use to see at GREC workshops, there are interesting problems that involve GR.

Graphics-rich document understanding

Documents following a graphical language (engineering drawings, architectural floor plans, maps, musical scores, diagrams, …)



Flowchart diagram recognition (e.g. patents)

CLEF-IP: retrieval in the intellectual property domain

http://www.ifs.tuwien.ac.at/~clef-ip/

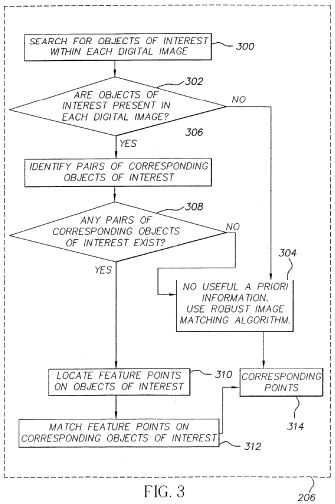
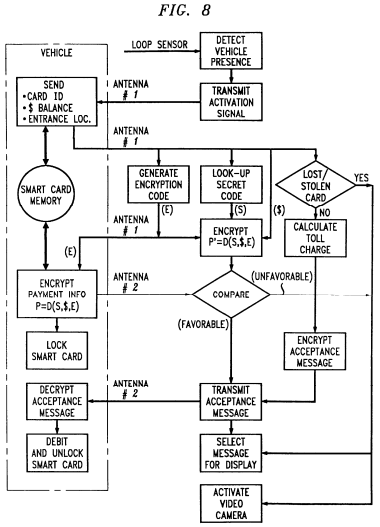
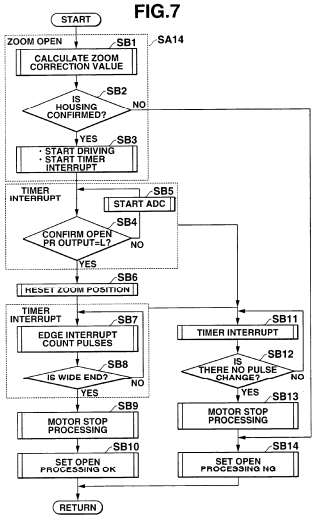
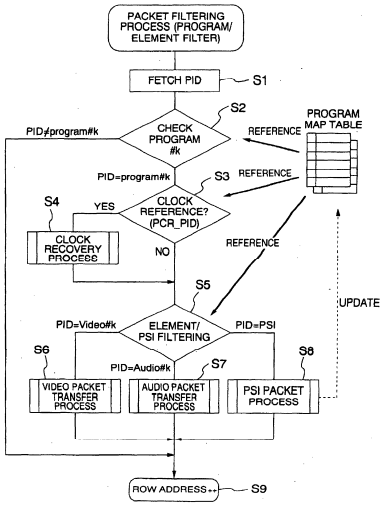
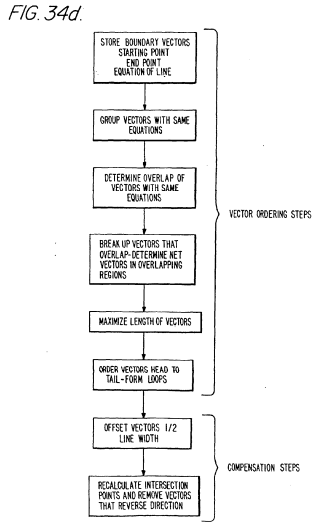


FIG. 3

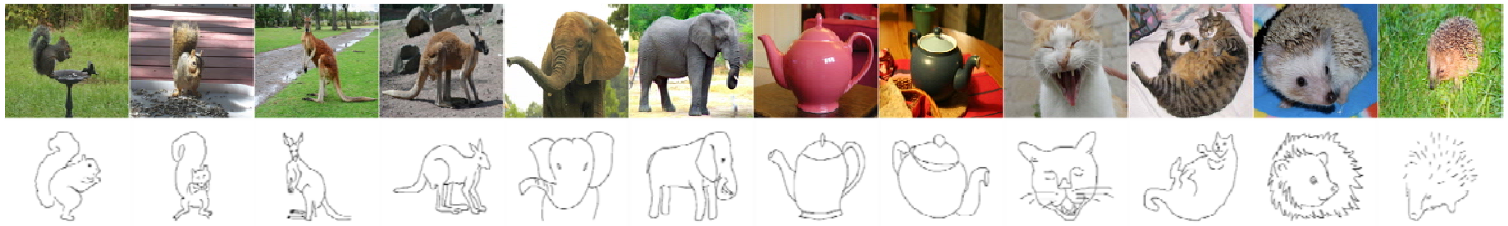






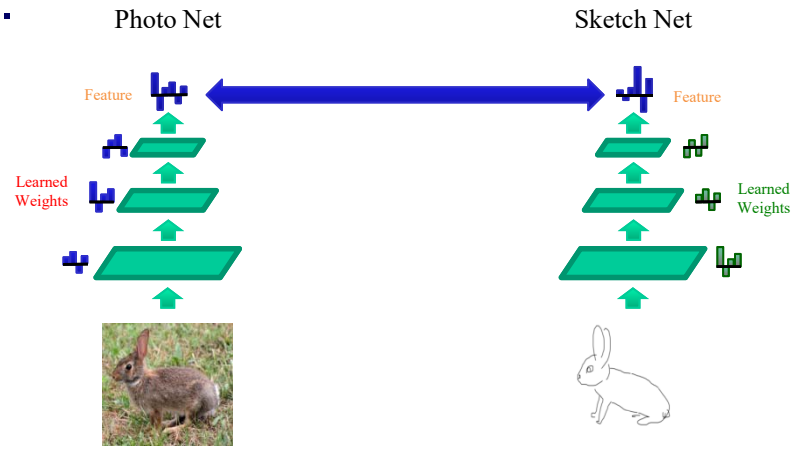


Sketch based image retrieval



- Sketchy database: large-scale collection of sketch-photo pairs.
- 125 categories: 75,471 sketches of 12,500 photographic objects.
- Used to train cross-domain convolutional networks which
- embed sketches and photographs in a common feature space.
- Benchmark for fine-grained retrieval.

Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. 2016. **The sketchy database: learning to retrieve badly drawn bunnies**. *ACM Trans. Graph.* 35, 4, Article 119 (July 2016), 12 pages. DOI: https://doi.org/10.1145/2897824.2925954



http://sketchy.eye.gatech.edu/

Sketched diagram understanding for X

Learning to Infer Graphics Programs from Hand-Drawn Images

Kevin Ellis
MIT
ellisk@mit.edu

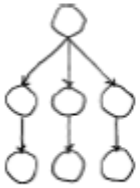
Daniel Ritchie
Brown University
daniel_ritchie@brown.edu

Armando Solar-Lezama
MIT
asolar@csail.mit.edu

Joshua B. Tenenbaum
MIT
jbt@mit.edu

Abstract

We introduce a model that learns to convert simple hand drawings into graphics programs written in a subset of LaTeX. The model combines techniques from deep learning and program synthesis. We learn a convolutional neural network that proposes plausible drawing primitives that explain an image. This set of drawing primitives is like an execution trace for a graphics program. From this trace we use program synthesis techniques to recover a graphics program with constructs such as variable bindings, iterative loops, or simple kinds of conditionals. With a graphics program in hand, we can correct errors made by the deep network, cluster drawings by use of similar high-level geometric structures, and extrapolate drawings. Taken together these results are a step towards agents that induce useful, human-readable programs from perceptual input.



```
Line(5,13, 2,10),
  arrow = True)
Circle(5,9)
Circle(8,5)
Line(2,8, 2,6),
  arrow = True)
Circle(2,5)
Circle(5,14)
Circle(2,9)
... etc. ...; 13 lines
```

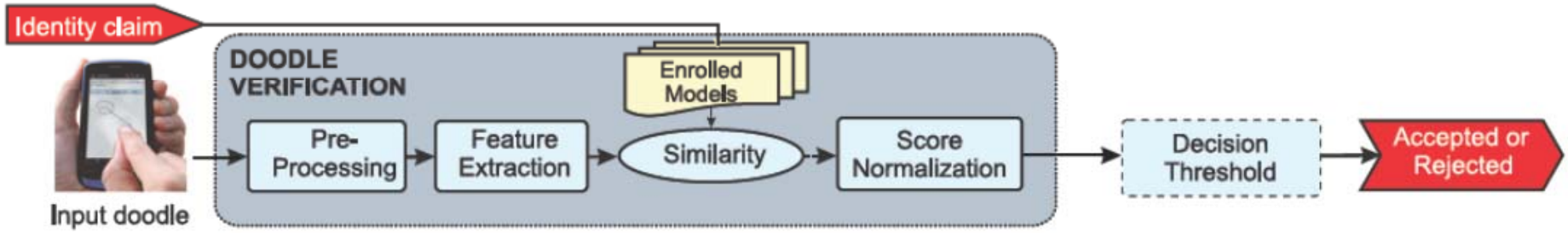
```
circle(4,10)
for(i<3)
 circle(-3*i+7,5)
 circle(-3*i+7,1)
 line(-3*i+7,4,-3*i+7,2,
      arrow=True)
 line(4,9,-3*i+7,6,
      arrow=True)
```

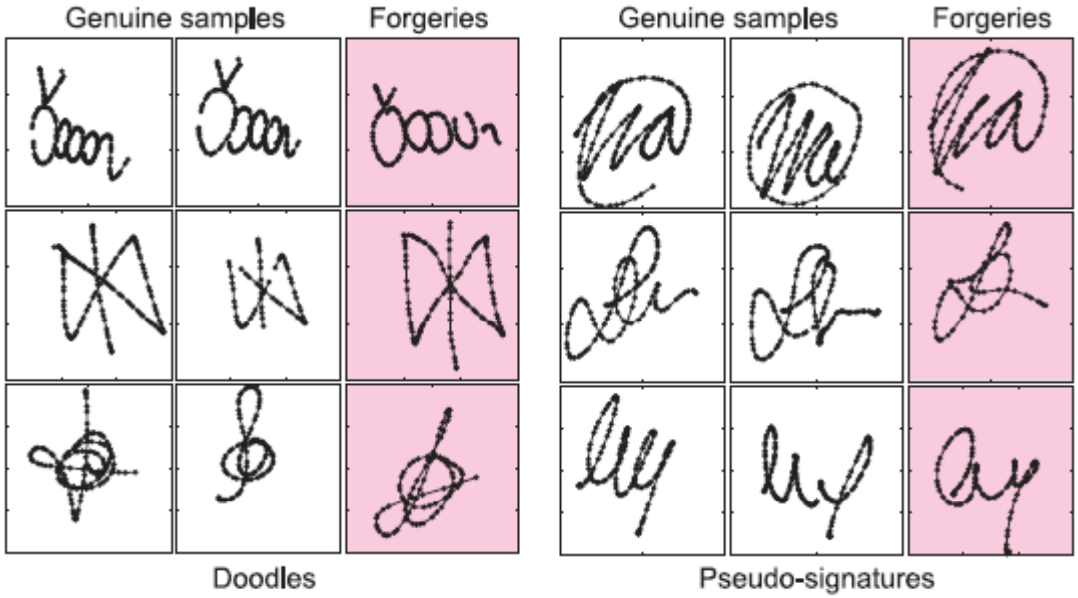
```
Circle(5,8)
Circle(2,8)
Circle(8,11)
Line(2,9, 2,10)
Circle(8,8)
Line(3,8, 4,8)
Line(3,11, 4,11)
Line(8,9, 8,10)
Circle(5,14)
... etc. ...; 21 lines
```

```
for(i<3)
 for(j<3)
  if(j>0)
   line(-3*j+8,-3*i+7,
        -3*j+9,-3*i+7)
  line(-3*i+7,-3*j+8,
       -3*i+7,-3*j+9)
  circle(-3*j+7,-3*i+7)
```

Graphical Passwords / Graphical User Authentification

A graphical password is an authentication system that works by having the user select from images, in a specific order, presented in a graphical user interface (GUI).





M. Martínez-Díaz, J. Fiérrez and J. Galbally Herrero.Graphical Password-Based User Authentication with Free-Form Doodles. IEEE Transactions on Human-Machine Systems 46.4 (2016): 607-614

Doodling expeeriences

- Google Quick Draw: https://quickdraw.withgoogle.com/#

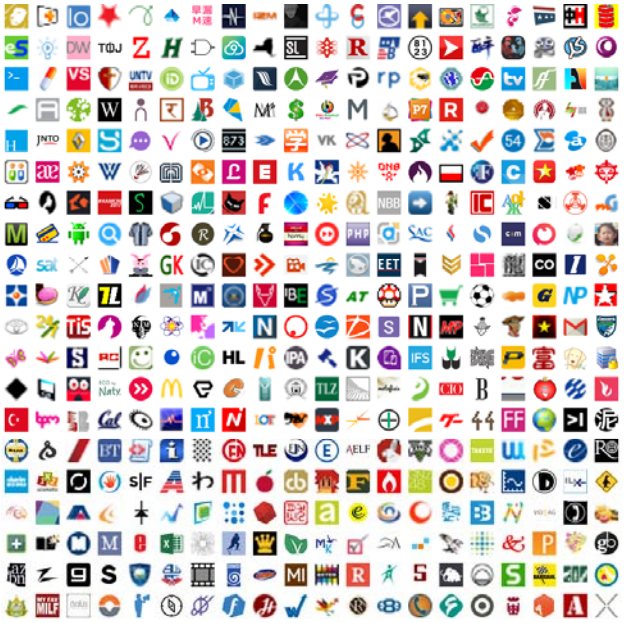
- Autodraw: https://www.autodraw.com/

Graffiti Recognition for Author Identification

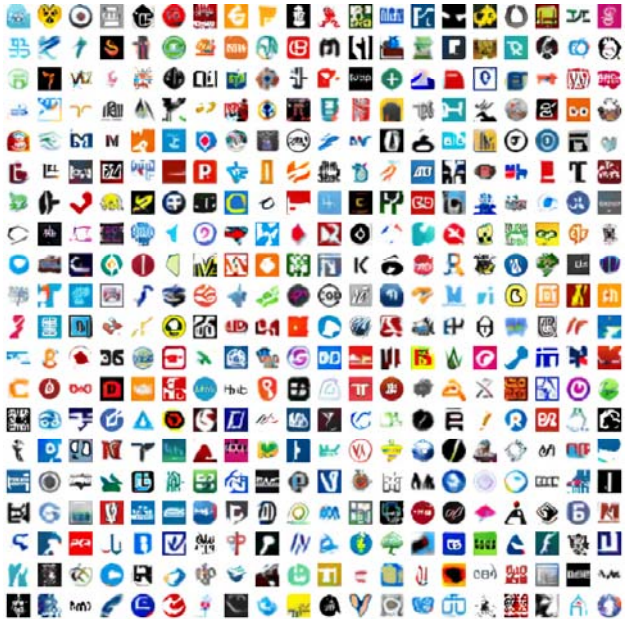
Logo Recognition

- New applications (e.g. in brand impact tracking in social media).
- Towards “scene graphics”
- Large Logo Dataset:

A. Sage, E. Agustsson, R. Timofte, and L.Van Gool, “Large Logo Dataset - version 0.1”, 2017



Real logos



Synthesized logos (with GANs)

GR in multimodal Question Answering

Are You Smarter Than A Sixth Grader?
Textbook Question Answering for Multimodal Machine Comprehension

Aniruddha Kembhavi† Minjoon Seo§* Dustin Schwenk† Jonghyun Choi†
Ali Farhadi†§ Hannaneh Hajishirzi§
†Allen Institute for Artificial Intelligence, §University of Washington
†{anik,dustins,jonghyunc,alif}@allenai.org, §{minjoon,hannaneh}@washington.edu

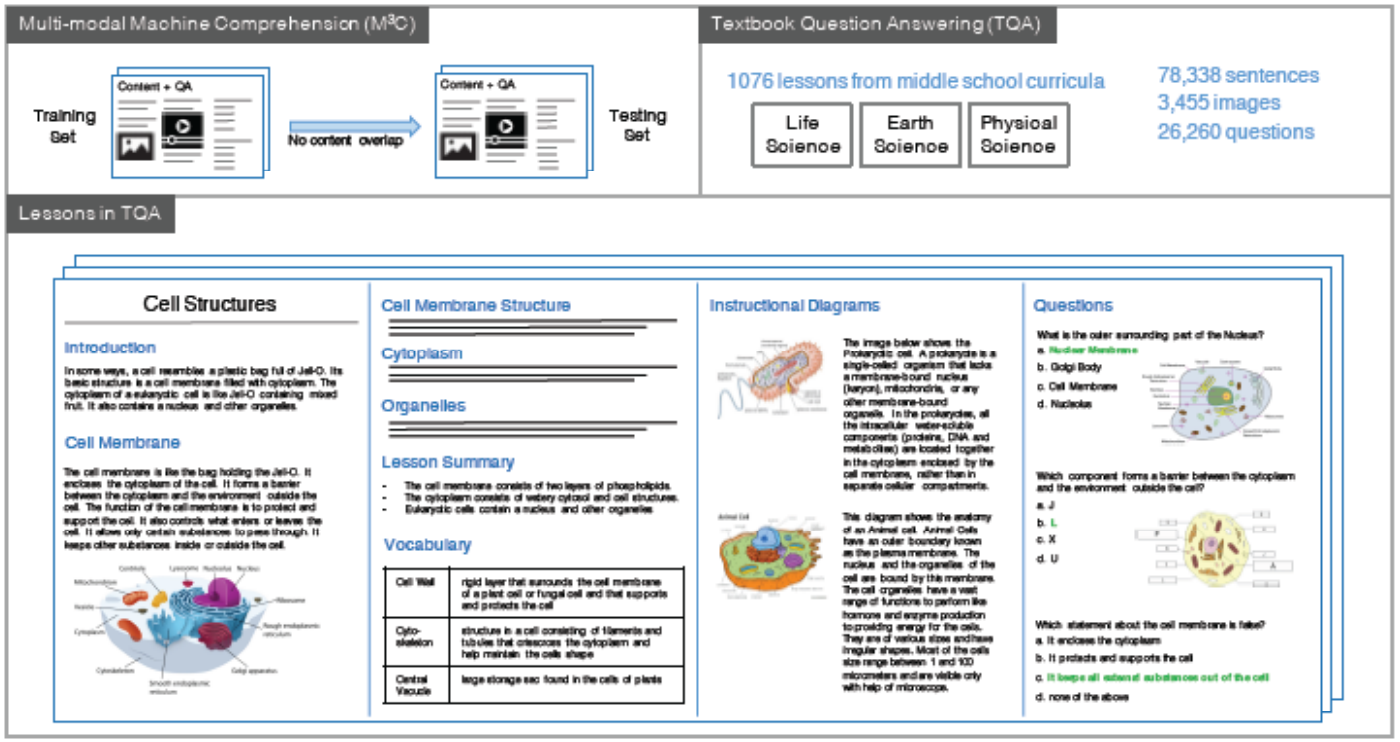
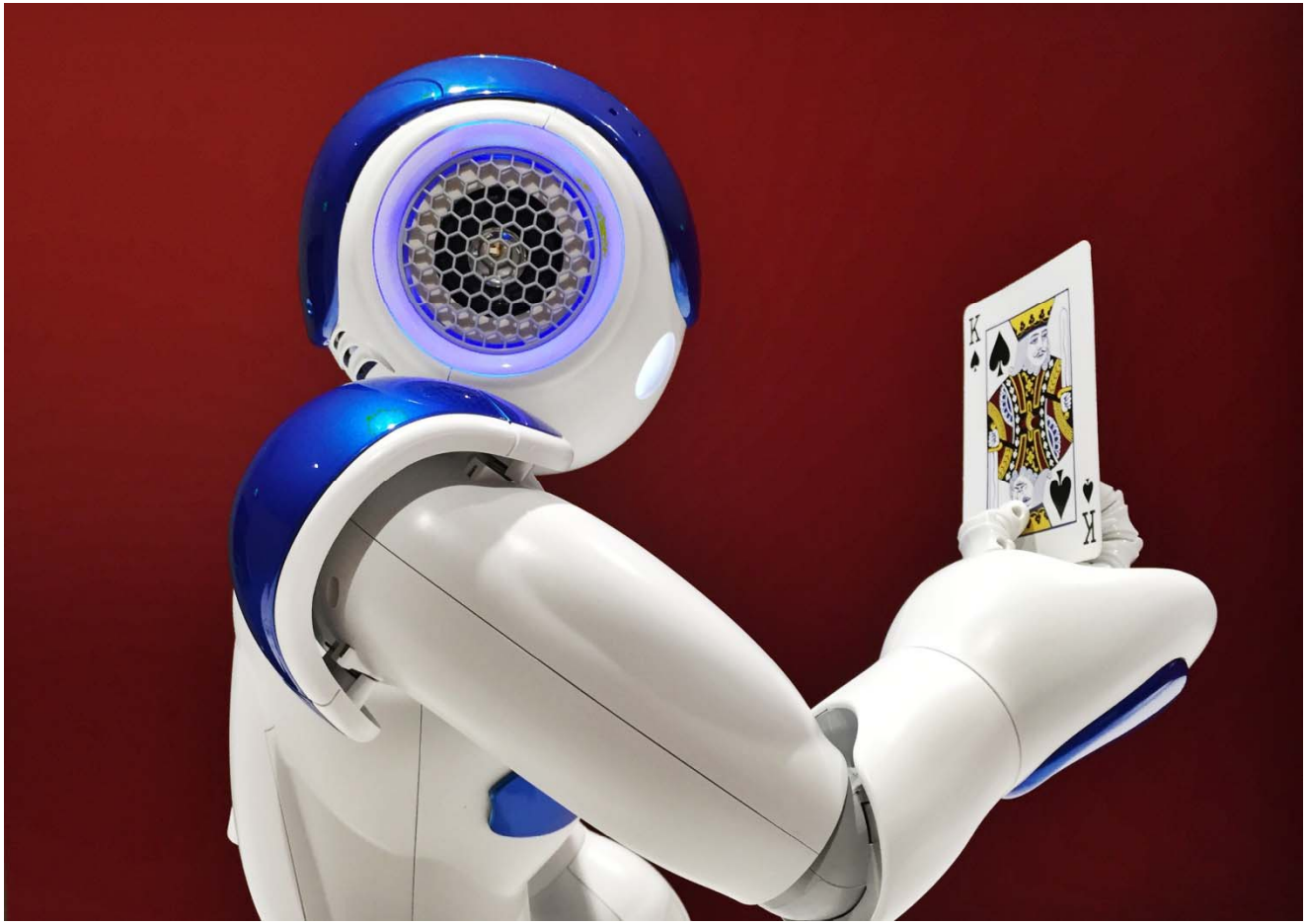


Figure 1. An overview of the Multi-modal Machine Comprehension (M^{3}C) paradigm, statistics of the proposed Textbook Question Answering (TQA) dataset and an illustration of a lesson in it. TQA can be downloaded at http://textbookqa.org .

IBM Watson's Visual Recognition Service for playing cards



And many more …

Final questions and challenges

- What is Graphics Recognition in 2017?
- Are we now more concerned in methodologies and apply them to Graphical Entities in end-to-end systems?
- Is GR an area by itself? Do we introduce ourselves as GR researchers? DIAR researchers? Where is the border between GR (TC10) and RS (TC11)?
- Which kind of annotated data do we need? How can we obtain it?
- What are the problems that GR can contribute to solve in the future? What is the society demanding? What are companies working on?
- Can somebody state our “great challenge”?
- Is anybody able to define the session topics of GREC’2027?