

Unveiling the Pulse of Computer Science¹

Paper reading session

Lei Wang^{1,2}

¹Australian National University

²Data61/CSIRO

September 7, 2023



Australian
National
University



¹This reading session covers (i) research trend in computer science, (ii) industry & academic research in computer vision and (iii) a closer look at citations.

Trends in Computer Science Research

European institutes provide major worldwide contributions to research in multiple CS domains².

- In 2020 alone, 1 million CS papers were published worldwide, $> 1/3$ by European researchers³.
- Leading domains e.g., algorithms and telecommunications, are \downarrow
- Uprising domains e.g., data science, and human-computer interaction, \uparrow .
- Some domains, e.g., computer vision, maintain $-$.

Table 1: Distribution of CS publications & citations within 10 years by Region

Region	Publications (%)	Citations (%)
Europe	$\sim 30\%$	$\sim 30\%$
North America	33%	$\sim 47\%$
Asia	30%	$\sim 16\%$
Other Continents	$< 6\%$	$< 6\%$

²Trends in Computer Science Research within European Countries, COMMUNICATIONS OF THE ACM, April 2022

³Microsoft Academic Graph (MAG) is a bibliometric dataset containing over 263M publications.

European Research and Collaborations

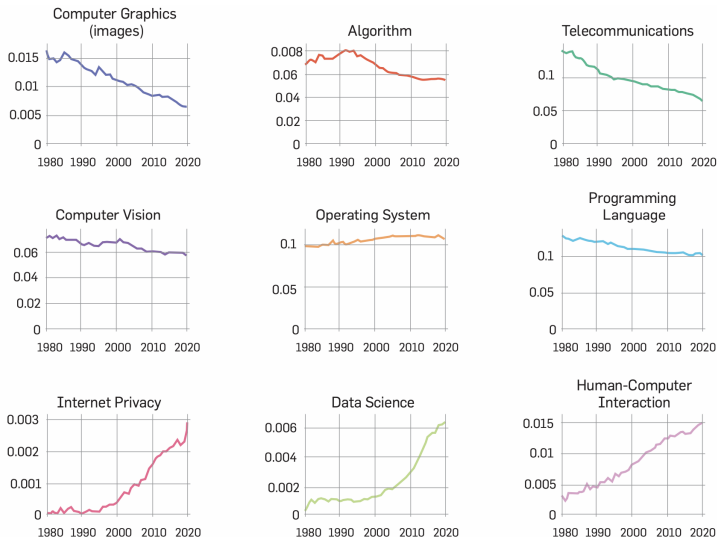


Figure 1: The number of European publications in different subfields, normalized by the total European CS publications.

Volume and Impact of Publications

Publication volumes are important, but do not necessarily imply quality or impact.

- Publication volume has been dramatically ↑ (*race to publish*)
- The spatial distribution of citations is similar to volume mapping.
 - Some smaller countries e.g., Switzerland & Netherlands of high impact
 - East-West impact gaps, possibly resulting from historical separation.
 - Switzerland, Italy, Netherlands, and Israel as **leading citation rates per institute**
 - Israel contributes **the highest institution citation rates in algorithms**, followed by Italy.
- Collaborations are essential for research
 - Russia & Poland exhibit the highest independent research
 - Belarus & Iceland, practice the highest ratios of collaborative research.
 - Some more collaborative countries (e.g., Iceland & Belarus) benefiting from greater citation impact **thanks to collaborations**

CS expands to Interdisciplinary Fields

- Growing interest related to health, energy, transportation, and aging.
- The highest European impacts are in health and public health, followed by energy related projects.
- Inspection of leading fields/countries/collaborations may support decision makers in distributing funds/shaping future (rapidly changing fields of CS).

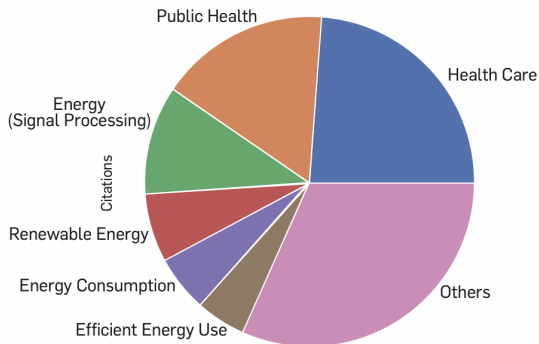


Figure 2: Percentage of CS papers in Europe related to the topics of health, energy, transportation, and aging.

Industry and Academic Research

Industry-sponsored & academic research have coexisted & interacted in all areas

- Industry:
 - motivated by short-term goals
 - aligned with business needs
 - specific product features
- Academic:
 - develop new models
 - improve existing models (SOTA)

Workflow differs

- Industry:
 - starts with fixed performance requirements
 - looks for a specific model/approach
 - satisfy them
 - scale up easily
- Where is my next great idea? *academia*
- More resources/data not available in the academia...

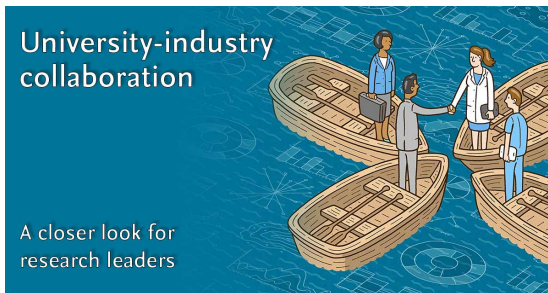
Dynamic between industry & academia

Research activity during the past decade: ↑

- Why growth?

- computational resources & large-scale data
- enormous value of computer vision apps → companies invest in R & D
- Results:
 - industry-sponsored researchers became a prominent part!
 - top CV conf. feature industry expos → opport. for networking & recruitment

But impact on CV research is unknown!



Some Informal Studies on Accepted Papers

- #papers in ICML where at least one author was affiliated with the industry increased from 20-25% in 2017 to 45% in 2018 ⁴.
- Publications by Google's DeepMind researchers alone doubled from 6% in ICML'17 to 13% in ICML'18 ⁵.
- Analysis of ICML and NeurIPS (in 2019) indicates the industry presence in 22.2% of papers^{6,7}
- Study of NeurIPS in 2014-2018 and ICML'17 and ICML'18 found that industry papers were among the most cited & many influential authors were collaborating with companies⁸.

⁴Andrej Karpathy, ICML accepted papers institution stats, 2017.

⁵Dhruv Guliani, ICML Accepted Papers Stats, 2018.

⁶Gleb Chuvpilo, AI Research Rankings 2019: Insights from NeurIPS and ICML, Leading AI Conferences, 2019.

⁷AndreasDoerr, ICML 2019 Accepted Paper Stats, 2019.

⁸Maithra Raghu, Citation Statistics of Machine Learning Papers, 2019.

Data Statistics

	Academia	Academia/Industry	Industry	Total
Number of papers	10,739	3,399	548	14,686
Number of authors	14,400	2,140	1,865	18,405
Number of citations	108,867	39,772	12,875	161,514
Number of affiliations	1,279	-	473	1,752
Number of code links	1,793	688	69	2,550

Figure 3: Dataset statistics⁹

- Publications accepted to major CV conferences from 2010 to 2019
- CVPR, ICCV, ECCV, ACCV, and BMVC were selected
- Meta-data:
 - authors & affiliations
 - affiliation type (industry/academia)
 - papers & corresponding year/venue/titles references/abstracts/topics

3Mformer: Multi-order Multi-mode Transformer for Skeletal Action Recognition

Lei Wang^{†,§} Piotr Koniusz^{*,§,†}
[†]Australian National University, [§]Data61/CSIRO
[§]firstname.lastname@data61.csiro.au

Uncertainty-DTW for Time Series and Sequences

Lei Wang^{*,†,§} and Piotr Koniusz^{*,§,†}
[†]Australian National University [§]Data61/CSIRO
[§]firstname.lastname@data61.csiro.au

Abstract. Dynamic Time Warping (DTW) is used for matching pairs of sequences and celebrated in applications such as forecasting the evolution of time

⁹Iuliia Kotseruba, Manos Papagelis, John K. Tsotsos, Industry and Academic Research in Computer Vision, 2021

Data Analysis & Visualizations

- #papers published only by industry-affiliated authors fluctuated
- #papers with authors both from academia & industry more than doubled
- papers with industry involvement comprise over 40% in 2019
- Most of new collaborations are between industry and academia.

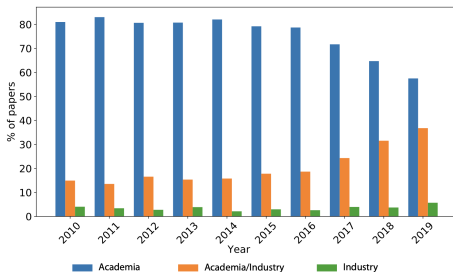


Figure 4: % of papers.

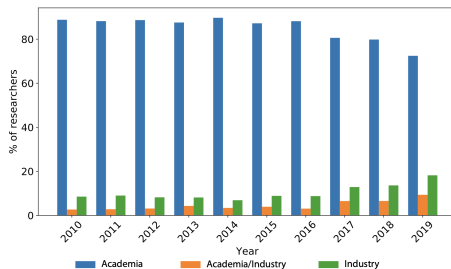
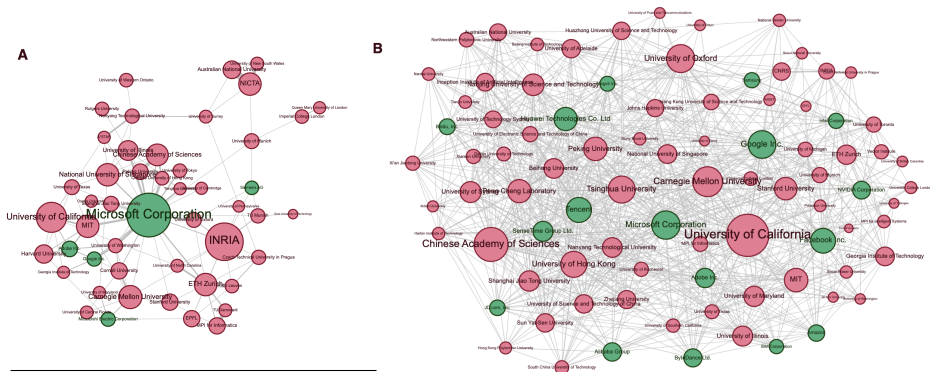


Figure 5: % of authors' affiliations.

- #researchers affiliated with industry has been growing steadily
- #authors with mixed affiliations increased from < 1% to 9%

Affiliation Network Visualization

- Part of the affiliation/collaborations network in (A) 2010 and (B) 2019
 - 2010: more discon./big connected component covering 60% of 439 nodes
 - 2019: #affiliations increased to 767/big component included 88% of the nodes
- Academia (red nodes) & industry (green nodes)¹⁰

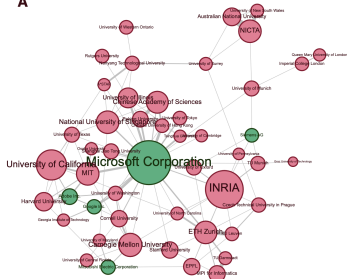


¹⁰Edges and nodes are weighted by the number of papers to which the affiliations contributed. In both networks, only nodes with node degree within top-10% are shown for clarity.

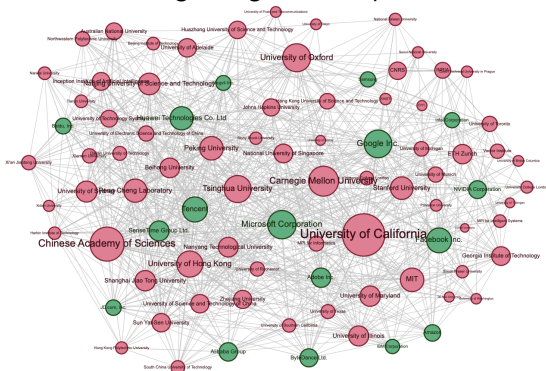
Affiliation Network Visualization (cont.)

- In 2010, only Microsoft was among the top-10 affiliations¹¹ & the rest were academic institutions.
- A decade later: 4 corporations (Google, Facebook, SenseTime, and Tencent) were among the top-10 contributors.
- Most of the top industry affiliations are fast-growing tech companies and start-ups¹²

A



B



¹¹Based on PageRank score

¹²Heavily rely on CV & ML for achieving business goals, e.g., marketing, social networks, and autonomous driving.

Affiliation Network Visualization (cont.)

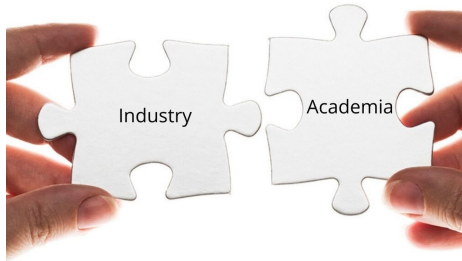
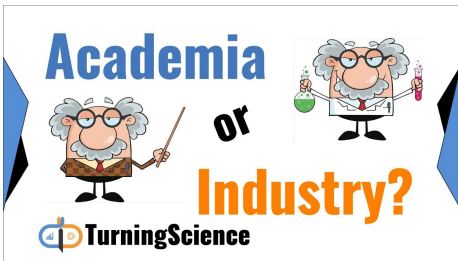
Connections between academic and industrial affiliations show regional patterns:

- Researchers from Google & Facebook often collaborate with top US universities e.g., CMU & UC
- Microsoft research divisions in Asia & Europe have strong connections with major schools in those regions (e.g., ETH Zurich in Switzerland & Tsinghua University in China).
- A number of Chinese corporations, e.g., Huawei/Tencent/SenseTime, collaborate with the largest



Impact of Industry Presence

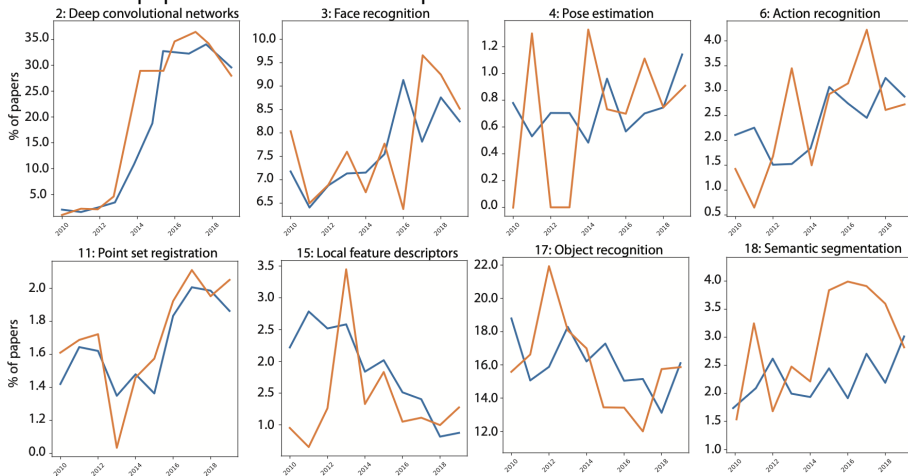
Companies offer better conditions/resources/monetary compensation for employees than most academic institutions



- Both recent graduates/established researchers may **shift research interests** towards topics that will increase the likelihood of **securing a job** in the industry.
- Some corporations also **offer various grants** to researchers and establish research labs within institutions.
 - Such grants & collaborations often go to researchers whose work is of **potential interest to the companies!**

Academic/Industry Research Topic Trends

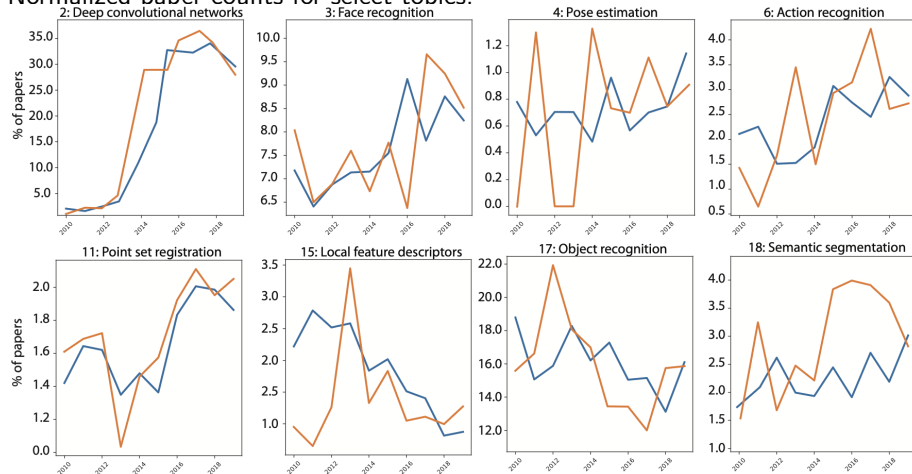
Normalized paper counts for select topics.



- Blue & red lines show academic & industry paper counts resp.
- Trends in research topics were evolving similarly in academia & industry

Academic/Industry Research Topic Trends (cont.)

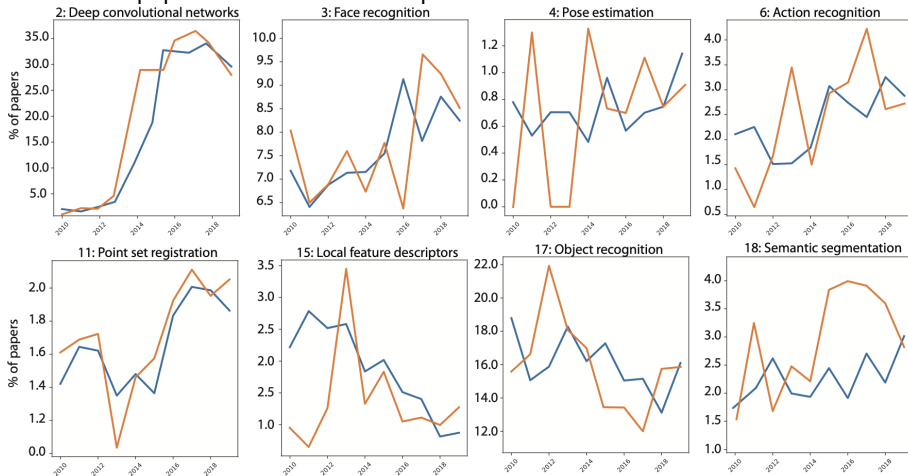
Normalized paper counts for select topics.



- Growing interest in DNN, #publications on face recognition/pose estimation/action recognition/point set registration ↑ (due to a large number of potential applications).

Academic/Industry Research Topic Trends (cont.)

Normalized paper counts for select topics.



- Handcrafted features ↓ (e.g., the advent of neural networks)
- Object recognition ↓ (e.g., benchmark results began to saturate)

Informal Studies on Citation¹⁴ Preferences

Papers produced by the industry have a larger influence on the community?

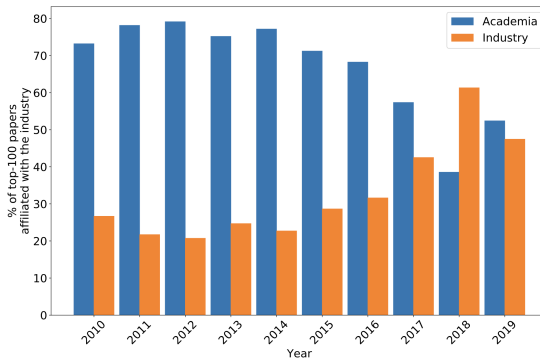
- **Industry papers are cited more frequently.** Out of top-10 cited papers:
 - 2 are from the industry
 - 3 are of mixed academic/industry authorship and
 - 5 are from the academic authors only
- **On average**, industry papers were cited 23 times vs. 11 for academia/industry & 10 for academic papers¹³.
- In the citation network
 - academic papers comprise 73%
 - only 4% of the papers were contributed by the industry
 - 23% by mixed academia/industry groups of authors
- **The 4% of industry papers gathered 8% of all citations!!!**

¹³Iuliia Kotseruba, Manos Papagelis, John K. Tsotsos, Industry and Academic Research in Computer Vision, 2021

¹⁴A citation is, by definition, an explicit declaration of influence by other papers, therefore, citation network is a reasonable source of data for inferring possible influences on researchers' choices. It should be noted that authors cite only a fraction of the papers that influenced them, and citation, in general, is subject to multiple biases, such as preference towards citing papers with already high citation count, papers with multiple co-authors, and publicly available implementations/data

A Closer Look at Citation Preferences

Proportion of academia & industry papers within top-100 papers (with highest PageRank score) for each year



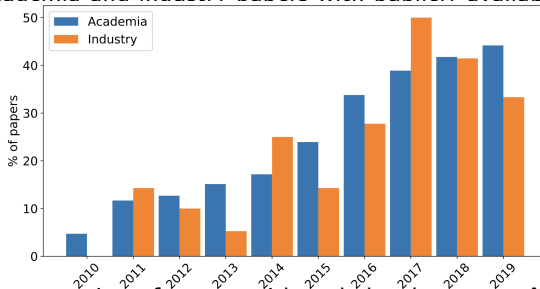
- Since 2015, industry-affiliated papers increased their influence
 - industry-affiliated papers overtook academic papers in 2018
 - nearly equal in influence to academic papers in 2019
- But 67/100 top papers over the whole decade were produced by academia

Potential Aspects of Citation Bias

Why citation bias towards industry papers???

The availability of code/data associated has been shown to give significant citation advantages in other areas of science¹⁵.

Proportion of academia and industry papers with publicly available code per year



- The relative proportion of papers with code has been **growing steadily**
- The proportion of publicly released code & the growth rate of the number of papers with code was \sim the same
- The influence of other papers¹⁶?

¹⁵Patrick Vandewalle, Code availability for image processing papers: a status update, 2019.

¹⁶When researching a topic, references from other similar papers are a common source of expanding bibliography.

Conclusion

This space
intentionally
left blank.

Acknowledgment: The plots/figures/statistics/tables presented in the slides are obtained from either online resources or newest publications (2021–2023).

Thank you!