Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Pan Zhang

New York university

Google Scholar profile

Generated 2026-06-10 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

748	749	30	73
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

87.5% independent of 726 classified citing papers

Citation type	Count
Independent	635
Self-citation	17
Co-author	74
Same-institution	0

23 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher advanced large multi-modal models by introducing ShareGPT4V, a framework that leverages improved captions to enhance model performance, establishing a significant benchmark in the field.

The researcher's primary contribution centers on the 2024 publication 'Sharegpt4v: Improving large multi-modal models with better captions.' This work appears to address the challenge of optimizing large multi-modal models by focusing on the quality of captioning data as a critical lever for improvement. The title suggests a methodological shift toward refining input data characteristics to drive model efficacy.

This line of work appears to fill a gap in understanding how caption quality directly impacts the capabilities of large multi-modal systems. By isolating 'better captions' as a variable for improvement, the researcher offers a targeted approach to model enhancement that differs from broader architectural changes. The absence of follow-up papers in the provided data indicates this core paper stands as a distinct, self-contained contribution to the literature.

The significance of this work is evidenced by its substantial citation count of 1195. Furthermore, citation analysis reveals that 87.5% of citing papers originate from independent researchers, indicating broad adoption and validation across the scientific community. This high degree of independent engagement suggests the work has become a foundational reference for subsequent studies in multi-modal learning.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 635 · **90 flagged influential by Semantic Scholar**

CORE PAPER

Sharegpt4v: Improving large multi-modal models with better captions

2024 · 1,195 citations (GS)

Field-normalised: 1,069 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2024 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Cambrian-1: A fully open, vision-centric exploration of multimodal llms	Korea Advanced Institute of Science and Technology (KAIST), Nanyang Technological University, New York University	Singapore, South Korea, United States	Influential
2	Instruction tuning for large language models: A survey	Alibaba Group, Nanyang Technological University, Peking University	China, Singapore, United States	Influential
3	Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning	SenseTime, Sensetime (China), SenseTime Research	Canada, China, Hong Kong	—
4	A survey on multimodal large language models	Nanjing University, Skywork AI, Tencent	China	Influential
5	Mm1. 5: Methods, analysis & insights from multimodal llm fine-tuning	Apple	United States	—
6	Dynamic-llava: Efficient multimodal large language models via dynamic vision-language context sparsification	East China Normal University, Nanjing University, Xiamen University	China	—

No.	Citing paper	Citing institution(s)	Country	S2
7	The synergy between data and multi-modal large language models: A survey from co-development perspective	Alibaba Group, Zhejiang University	China	—
8	Imp: Highly capable large multimodal models for mobile devices	Hangzhou Dianzi University	China	—
9	Efficiently integrate large language models with visual perception: A survey from the training paradigm perspective	Lingnan University	Hong Kong	—
10	Poison as cure: Visual noise for mitigating object hallucinations in lvms	Alibaba Group, Westlake University, Xiamen University	China	—
11	Lanp: Rethinking the impact of language priors in large vision-language models	Chongqing University, Peking University, Penn State University	China, Singapore, United States	—
12	Modality curation: Building universal embeddings for advanced multimodal information retrieval	Kuaishou Technology, Northeastern University	China, United States	—
13	A survey of multimodal retrieval-augmented generation	Huawei Technologies (China)	China	—
14	Vision-flan: Scaling human-labeled tasks in visual instruction tuning	University of Illinois Urbana-Champaign, University of Michigan, University of Washington	United States	—
15	Forensics-bench: A comprehensive forgery detection benchmark suite for large vision language models	Alibaba, MEGVII Technology, Shanghai AI Laboratory	China, Hong Kong	**Influential**
16	Benchmarking multimodal large language models for face recognition	Idiap Research Institute	Switzerland	—
17	Muse-vl: Modeling unified vlm through semantic discrete encoding	ByteDance	China	—
18	Simulating the real world: A unified survey of multimodal generative models	Bosch (China) Investment Co., Ltd., PLA Air Force Aviation University, The Chinese University of Hong Kong	China	—
19	Multimodal foundation models: From specialists to general-purpose assistants	Microsoft, Microsoft Corporation, National University of Singapore	Singapore, United States	—
20	Mammoth-vl: Eliciting multimodal reasoning with instruction tuning at scale	Beijing University of Posts and Telecommunications, Carnegie Mellon University, Nanyang Technological University	Canada, China, Singapore	—
21	Janusflow: Harmonizing autoregression and rectified flow for unified multimodal understanding and generation	DeepSeek-AI, Peking University, ShanghaiTech University	China, Hong Kong	—
22	Blink: Multimodal large language models can see but not perceive	Columbia University, University of Pennsylvania, University of Washington	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
23	Sft memorizes, rl generalizes: A comparative study of foundation model post-training	New York University, The University of Hong Kong, UC Berkeley	Hong Kong, United States	—
24	Mm1: methods, analysis and insights from multimodal llm pre-training	Apple, Microsoft	United States	—
25	Show-o: One single transformer to unify multimodal understanding and generation	ByteDance, National University of Singapore	China, Singapore	Influential
26	What matters when building vision-language models?	Hugging Face, Sorbonne Université	France, United States	—
27	Janus: Decoupling visual encoding for unified multimodal understanding and generation	DeepSeek-AI, Peking University, The University of Hong Kong	China, Hong Kong	—
28	Show-o2: Improved native unified multimodal models	Johns Hopkins University, National University of Singapore	Singapore, United States	—
29	Vila-u: a unified foundation model integrating visual understanding and generation	MIT, NVIDIA, Tsinghua University	China, United States	—
30	Tokenflow: Unified image tokenizer for multimodal understanding and generation	ByteDance	China	—

Showing the 30 most-cited of 635 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 2

Claim — Contribution 2

The researcher advanced spectral methods for clustering sparse networks, establishing a foundational framework that has been widely adopted by independent scholars.

The researcher's core contribution rests on the 2013 paper 'Spectral redemption in clustering sparse networks,' which appears to address challenges in applying spectral techniques to sparse network structures. This work suggests a methodological refinement or theoretical correction, indicated by the term 'redemption,' aimed at improving clustering accuracy in sparse regimes.

The originality of this line of work lies in its focus on sparse networks, a domain where traditional spectral methods often struggle. By targeting this specific gap, the researcher provided a specialized solution that likely enhanced the robustness or applicability of spectral clustering algorithms for sparse data.

The significance of this contribution is evidenced by its high citation count of 824. Notably, 87.5% of the citing papers originate from independent researchers, indicating that the work has been broadly adopted and validated by the wider scientific community rather than just the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

Spectral redemption in clustering sparse networks

2013 · 824 citations (GS)

Field-normalised: 653 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2013 indexed by Semantic Scholar, by citation count.

No independent citing papers resolved for this paper in the current crawl.

Contribution 3

Claim — Contribution 3

The researcher critically re-evaluates standard assessment protocols for large vision-language models, establishing a foundational framework that has garnered significant independent scholarly attention.

CLAIM: The researcher's primary contribution is a critical examination of evaluation methodologies for large vision-language models, anchored by the 2024 paper titled 'Are we on the right way for evaluating large vision-language models?' This work serves as the cornerstone of this specific line of inquiry.

ORIGINALITY: The title suggests a pivotal shift in perspective, questioning established norms rather than merely proposing incremental improvements. By challenging the status quo of model assessment, the researcher appears to address a fundamental gap in how the community validates the capabilities and limitations of these complex systems.

SIGNIFICANCE: The work has achieved substantial impact, evidenced by 902 citations. Notably, 87.5% of the classified citing papers originate from independent researchers, indicating that the contribution has resonated broadly across the field and influenced external scholarly discourse beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

Are we on the right way for evaluating large vision-language models?

2024 · 902 citations (GS)

Field-normalised: 738 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2024 indexed by Semantic Scholar, by citation count.

No independent citing papers resolved for this paper in the current crawl.

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	65
The Chinese University of Hong Kong	Hong Kong	SCImago #163 · THE =41 · QS =32	58
Shanghai AI Laboratory	China	—	56
Shanghai Jiao Tong University	China	SCImago #10 · THE 40 · QS =47	56
Peking University	China	SCImago #11 · THE 13 · QS 14	53
University of Science and Technology of China	China	SCImago #77 · THE 51 · QS =132	52
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	43
Nanyang Technological University	Singapore	SCImago #137	42
Nanjing University	China	SCImago #178 · THE =62 · QS =103	37
Shanghai Artificial Intelligence Laboratory	China	SCImago #563	35

Institution	Country	World ranking	Citing papers
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	33
The University of Hong Kong	Hong Kong	SCImago #195 · THE 33 · QS 11	30
Alibaba Group	China	SCImago #226	29
Fudan University	China	SCImago #46 · THE 36 · QS 30	29
ByteDance	China	—	28

Geographic distribution of citing authors

Country	Citing papers
China	508
United States	210
Hong Kong	80
Singapore	76
United Kingdom	38
South Korea	28
Australia	25
Canada	25
Germany	15
United Arab Emirates	12
Japan	12
France	9

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).

- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Sharegpt4v: Improving large multi-modal models with better captions	635	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Spectral redemption in clustering sparse networks	0	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Are we on the right way for evaluating large vision-language models?	0	8 CFR 204.5(h)(3)(v) — Criterion 5