Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Zirui Wang

University of Virginia

Google Scholar profile

Generated 2026-05-21 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

91	92	6	3
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.

92.3% independent of 52 classified citing papers

Citation type	Count
Independent	48
Self-citation	2
Co-author	2
Same-institution	0

39 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 developed a low-rank audio-visual fusion method guided by temporal cues to improve the accuracy of automated video highlight detection systems.

The researcher's core contribution rests on the 2021 paper 'Temporal cue guided video highlight detection with low-rank audio-visual fusion.' This work appears to introduce a specific technical approach for identifying salient moments in video content by integrating audio and visual data through low-rank fusion techniques, guided by temporal cues. The titles indicate a focus on multimodal signal processing to enhance detection precision.

This line of work addresses the challenge of effectively synchronizing and fusing heterogeneous audio-visual streams for video analysis. By leveraging temporal cues and low-rank structures, the research suggests a novel way to reduce noise and redundancy in multimodal data, potentially offering a more robust solution than unimodal or loosely coupled approaches. The absence of follow-up papers by the same researcher in the provided data means this core publication stands as the primary artifact of this specific methodological contribution.

The significance of this work is evidenced by its citation record, with 79 citations indicating substantial uptake in the field. Notably, 92.3% of the classified citing papers originate from independent researchers, suggesting that the methodology has been adopted and built upon by the broader academic community rather than just the researcher's immediate circle. This high degree of independent citation underscores the work's impact and utility in advancing video highlight detection research.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 43 · **3 flagged influential by Semantic Scholar**

CORE PAPER

Temporal cue guided video highlight detection with low-rank audio-visual fusion

2021 · Proceedings of the IEEE/CVF International Conference on Computer Vision ..., 2021 · 79 citations (GS)

Field-normalised: 58 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2021 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	A survey of multimodal learning: Methods, applications, and future	Northwestern Polytechnical University	China	—
2	Univtg: Towards unified video-language temporal grounding	Johns Hopkins University, Meta AI, National University of Singapore	Singapore, United States	—
3	Query-dependent video representation for moment retrieval and highlight detection	Sungkyunkwan University	South Korea	—
4	Video summarization techniques: A comprehensive review	Egypt-Japan University of Science and Technology, Egypt-Japan University of Science and Technology; Alexandria University	Egypt	**Influential**
5	Tr-detr: Task-reciprocal transformer for joint moment retrieval and highlight detection	Central China Normal University, Hubei University of Technology	China	—
6	Correlation-guided query-dependency calibration for video temporal grounding	Sungkyunkwan University	South Korea	—

No.	Citing paper	Citing institution(s)	Country	S2
7	Umt: Unified multi-modal transformers for joint video moment retrieval and highlight detection	Tencent PCG, The Hong Kong Polytechnic University	China, Hong Kong	Background
8	Learning content-enhanced mask transformer for domain generalized urban-scene segmentation	University of Amsterdam	Netherlands	—
9	Task-driven exploration: Decoupling and inter-task feedback for joint moment retrieval and highlight detection	Xi'an Jiaotong University	China	—
10	-Tuning: Efficient Image-to-Video Transfer Learning for Video Temporal Grounding	The Hong Kong Polytechnic University	China	—
11	Prior knowledge integration via llm encoding and pseudo event regulation for video moment retrieval	The Hong Kong Polytechnic University	China, Hong Kong	—
12	Hyperbolic audio-visual zero-shot learning	Data61-CSIRO, Monash University, Southeast University	Australia, China	Background
13	Spikemba: Multi-modal spiking saliency mamba for temporal video grounding	Harbin Institute of Technology	China	—
14	Flashvtg: Feature layering and adaptive score handling network for video temporal grounding	The University of Queensland	Australia	—
15	Counterfactual cross-modality reasoning for weakly supervised video moment localization	Renmin University of China	China	Background
16	Moment Quantization for Video Temporal Grounding	Dolby Laboratories, Xi'an Jiaotong University	China, United States	—
17	Cav-mae sync: Improving contrastive audio-visual mask autoencoders via fine-grained alignment	Goethe University of Frankfurt, IBM Research, IBM Research, MIT-IBM Watson AI Lab	Germany, Japan, United States	—
18	Ms-detr: Towards effective video moment retrieval and highlight detection by joint motion-semantic learning	Fudan University, Tencent	China	—
19	Cfsum: A transformer-based multi-modal video summarization framework with coarse-fine fusion	University of California, Los Angeles, Zhejiang University	China, United States	—
20	Unsupervised modality-transferable video highlight detection with representation activation sequence learning	Tencent	China	**Methodology**
21	Which Viewpoint Shows it Best? Language for Weakly Supervising View Selection in Multi-view Instructional Videos	FAIR, Meta, University of Texas at Austin, University of Utah	France, United States	—
22	DiffusionVMR: Diffusion model for joint video moment retrieval and highlight detection	Nanjing University, Nanjing University of Science and Technology, University of Washington	China, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
23	Actprompt: In-domain feature adaptation via action cues for video temporal grounding	Microsoft Research Asia, Tongji University, Xidian University	China	—
24	Query-centric audio-visual cognition network for moment retrieval, segmentation and step-captioning	Beihang University, University of Chinese Academy of Sciences	China	—
25	Unsupervised video highlight detection by learning from audio and visual recurrence	Google DeepMind, University of Saskatchewan	Canada, United Kingdom	—
26	Which viewpoint shows it best? language for weakly supervising view selection in multi-view instructional videos	FAIR, Meta, University of Texas at Austin, University of Utah	France, United States	—
27	Ld-detr: Loop decoder detection transformer for video moment retrieval and highlight detection	Sun Yat-sen University	China	—
28	Test-Time Adaptation for Video Highlight Detection Using Meta-Auxiliary Learning and Cross-Modality Hallucinations	Google DeepMind, University of Saskatchewan	Canada, United Kingdom	—
29	Finding the right moment: Human-assisted trailer creation via task composition	University of Edinburgh	United Kingdom	—
30	Probing visual-audio representation for video highlight detection via hard-pairs guided contrastive learning	SenseTime Research	China	Methodology

Showing the 30 most-cited of 43 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).

Citing-text excerpts — how the field used this work

METHODOLOGY Unsupervised modality-transferable video highlight detection with representation activation sequence learning

"References [9] and [10] need both the visual and the audio modalities as the input in the training and inference processes."

METHODOLOGY Probing visual-audio representation for video highlight detection via hard-pairs guided contrastive learning

"We compare our proposed method with the other state-of-the-art methods [1, 13, 17, 32, 39, 40, 41, 43, 44] on two widely adopted benchmarks, i . e ., YouTube Highlights and TVSum."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	4
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	3
Peking University	China	SCImago #11 · THE 13 · QS 14	3
Alibaba Group	China	SCImago #226	3
The Hong Kong Polytechnic University	Hong Kong	SCImago #256 · THE 80 · QS 54	3
Sungkyunkwan University	South Korea	SCImago #527 · THE 87 · QS =126	3

Institution	Country	World ranking	Citing papers
Nanjing University of Science and Technology	China	SCImago #541 · THE 601–800 · QS 701-710	2
University of Edinburgh	United Kingdom	SCImago #182 · THE 29 · QS 34	2
Guru Ghasidas Vishwavidyalaya	India	SCImago #8274 · THE 1501+	2
University of Utah	United States	SCImago #320 · THE 201–250 · QS =540	2
Georgia Institute of Technology	United States	SCImago #270 · THE =41 · QS =123	2
DAMO Academy, Alibaba Group	United States	—	2
Renmin University of China	China	SCImago #2319	2
UT Austin	United States	—	2
University of Amsterdam	Netherlands	SCImago #75 · THE =62 · QS 53	2

Geographic distribution of citing authors

Country	Citing papers
China	31
United States	14
Singapore	4
United Kingdom	4
India	3
South Korea	3
Australia	2
Canada	2
France	2
Hong Kong	2
Netherlands	2
Saudi Arabia	1

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	Temporal cue guided video highlight detection with low-rank audio-visual fusion	43	8 CFR 204.5(h)(3)(v) — Criterion 5