

Multimodal vs. Text-Only Similarity Measures in Web Content Clustering Performance

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How do multimodal similarity measures (e.g., CLIP embeddings) compare to text-only measures (e.g., SimGCL, DCL) in clustering web-page content that includes both text and images, as measured by. We propose a dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network for the challenging task of brain lesion segmentation. The devised architecture is the result of an in-depth analysis of the limitations of current networks proposed for similar. 11 claims were extracted from source literature; 8 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Research question: How do multimodal similarity measures (e.g., CLIP embeddings) compare to text-only measures (e.g., SimGCL, DCL) in clustering web-page content that includes both text and images, as measured by normalized mutual information (NMI) and F1-score on a benchmark dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.4/10.

3 Results

12 papers retrieved. 11 claims extracted; 8 independently verified. Quality review score: 7.4/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The proposed architecture is a dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network.	✓	0.25
The method employs a dense training scheme that joins the processing of adjacent image patches into one pass through the	✓	0.23
The training scheme automatically adapts to the inherent class imbalance present in the data.	✓	0.16
The architecture uses a dual pathway design to process input images at multiple scales simultaneously.	✓	0.16
A 3D fully connected Conditional Random Field is used for post-processing the network's soft segmentation.	✓	0.27
The 3D fully connected Conditional Random Field effectively removes false positives.	✓	0.26
The pipeline was evaluated on multi-channel MRI patient data for traumatic brain injuries, brain tumours, and ischemic s	✓	0.27
The method achieved top ranking performance on the BRATS 2015 public benchmark.	×	0.11
The method achieved top ranking performance on the ISLES 2015 public benchmark.	×	0.11
The method improves on the state-of-the-art for lesion segmentation in traumatic brain injuries, brain tumours, and isch	✓	0.24
The source code of the implementation is made publicly available.	×	0.12

References

- <https://doi.org/10.1186/s40537-021-00492-0>
- <https://doi.org/10.1007/s11263-016-0981-7>
- <https://doi.org/10.1016/j.media.2016.10.004>