Responsible NLP Checklist

Paper title: *RAV: Retrieval-Augmented Voting for Tactile Descriptions Without Training*
Authors: *Jinlin Wang, Yulong Ji, Hongyu Yang*

How to read the checklist symbols:

☑ the authors responded 'yes'

☒ the authors responded 'no'

N/A the authors indicated that the question does not apply to their work

☐ the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the Responsible NLP Checklist page at ACL Rolling Review.

☑ **A. Questions mandatory for all submissions.**

☑ A1. Did you describe the limitations of your work?
This paper has a Limitations section.

☒ A2. Did you discuss any potential risks of your work?
Our work introduces a foundational method (RAV) for generating tactile descriptions. The 'Limitations' section already discusses factors affecting description accuracy and reliability. These are primarily performance limitations, not direct, foreseeable societal risks or misuse potentials inherent to the RAV method itself. Broader risks would depend on specific downstream applications, which are outside this paper's scope.

☑ **B. Did you use or create scientific artifacts? (e.g. code, datasets, models)**

☑ B1. Did you cite the creators of artifacts you used?
We have diligently cited the creators of all artifacts used in our research. This includes citations for the datasets (TVL, SSVTP, HCT), the foundational and comparative models (e.g., CLIP, GPT-4V, LLaVA), and the key software libraries used for evaluation. All corresponding papers and sources are listed in the References section.

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
Created: Implementation code for our RAV method (including SyncVote, DualVote, WeightVote), constructed visual (My) and tactile (MT) vector databases, model checkpoints, and experimental/evaluation scripts. Used: CLIP model, TVL dataset (SSVTP, HCT), and GPT-4V (for evaluation). We will open-source our code and checkpoints at https://github.com/PluteW/RAV.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Our use of all existing artifacts, such as the TVL dataset and CLIP model, is consistent with their specified intended use for academic research. For our created artifact, the RAV framework, we specify its intended use is for research and educational purposes in multimodal perception and robotics. Crucially, as RAV is a training-free method, our released code (Apache 2.0) does not contain derivatives of the datasets (like trained weights). Users must obtain the datasets themselves,

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice.

thereby agreeing to their original access conditions (e.g., for research use only), which ensures compatibility and prevents misuse.

☒ B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?
We used the publicly available TVL dataset, which was collected and anonymized by its creators. The data primarily consists of images of material textures from robotic. The dataset contains images of interacting with objects, but no other personally identifiable information (PII) such as faces or names is present. The textual labels are from a controlled, non-offensive vocabulary of tactile properties. We reviewed the data and confirmed it contains no direct PII or offensive content, so no further anonymization or filtering steps were required on our part.

N/A B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
(left blank)

☑ B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?
Section 4.1 Dataset.

☑ **C. Did you run computational experiments?**

☒ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Our work primarily introduces a foundational methodology, RAV (Retrieval-Augmented Voting), for generating tactile descriptions from visual-tactile data. The core of this method lies in its parameter-free knowledge construction and voting mechanism.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4, Experiment and Result.

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Section 4.2, Tactile Description

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
Section 4.2, Tactile Description

☒ **D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

N/A D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
(left blank)

N/A D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
(left blank)

☑ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?
Section 4.1, Dataset

☒ N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
(left blank)

☒ N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
(left blank)

☑ **E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?**

☒ E1. If you used AI assistants, did you include information about their use?
AI tools were used for auxiliary proofreading to check for errors in the initial draft of the paper. The core writing, subsequent revisions, and finalization were entirely performed by the authors, with the use of AI tools being strictly controlled and limited to this proofreading assistance.