X-TASAR: An Explainable Token-Selection Transformer Approach for Arabic Sign Language Alphabet Recognition

Anonymous Author(s)
Affiliation
Address
email

Abstract

We propose a multistage transformer-based architecture for efficient Arabic Sign Language (ArSL) recognition. The proposed approach first extracts a compact 7×7 grid of image features using a tiny Swin transformer. We next determine a class-conditioned score of each grid token with the query [CLS] and pick a diverse Top-K subset through grid non-maximum suppression (NMS) algorithm. Only these K selected tokens together with [CLS] are then subjected to a small transformer-based classifier (ViT Tiny) to obtain the final label. The colored heatmap in the visualizations indicates which sections of the images had the highest scores, and the dots indicate the exact patches the classifier relied on to make its decision. Our model achieves 98.1% accuracy and 0.979 macro-F1 on the held-out test split on the RGB ArSl alphabet dataset (32 classes, 54049 images of more than xx signers each). It is also computationally lighter than a ViT-Tiny baseline as it reads only K+1 tokens instead of all 196 patches. The proposed approach is backbone-agnostic and can be adapted into other vision transformers with minimal modification, enabling accessible and scalable sign-language recognition tools for Arabic-speaking deaf and hard-of-hearing communities worldwide.

1 Introduction

Arabic Sign Language Recognition (ArSLR) is most commonly used in assistive technologies such as in teaching, government services, and human-computer interaction in Arabic-speaking communities [1]. Real-world implementation is characterized by unique challenges such as high signer variability (hand size, speed, habit), background interference (home/classroom environments), variations in illumination, and handshapes [8]. In the past, isolated-sign recognition has evolved past handcrafted descriptors (skin-color heuristics, HOG/SIFT-like features) toward CNNs trained on RGB images, and continuous signing frequently addressed by temporal models (HMMs, CRF, LSTM/GRU) [4]. More recently, vision transformers (ViT, Swin) achieved higher accuracy but at higher computational cost; parallel directions include pose/skeleton cues (2D/3D keypoints) or cross-modal models (CLIP-style adaptations) [3]. Two practical gaps remain: (1) efficiency — models that attend to all patches slow down rapidly as images have more patches; and (2) clear explanations — it's often unclear which parts of the image actually drove the prediction.

In this study, we propose X-TASAR (E*X*plainable *T*oken-selection transformer approach for *A*rabic *S*ign language *A*lphabet *R*ecognition), an explanation-driven and computationally efficient method. Our method extracts a compact 7×7 grid of image features using a tiny Swin transformer [6], adds a class-conditioned score to each of the grid tokens, and picks a diverse Top-K subset using non-maximum suppression (NMS) algorithm [7]. To perform the final classification, a small global ViT-like transformer [5] considers the chosen K tokens and [CLS], to predict the final classification

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

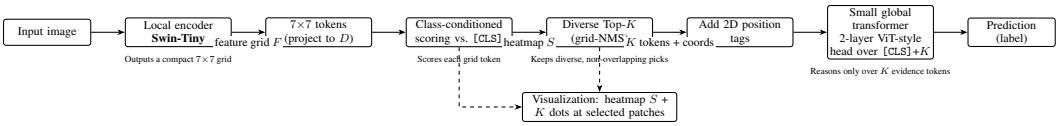

Figure 1: X-TASAR uses a tiny Swin encoder $7\times7\times768$), a linear projection to D=384, and a class-conditioned scorer. We select a diverse set of K=16 tokens with grid-NMS (radius r=1) and pass [CLS]+K tokens to a smaall ViT-style head with 2 encoder blocks and 6 attention heads.

label. The same attention map that drives the final prediction is visualized as a colored heatmap along with dots mark indicating the exact patches forwarded to the classifier, yielding simple, readable overlays.

2 Method

2.1 Overview

X-TASAR is a three-stage transformer pipeline for isolated ArSLR (Fig. 1): (i) a compact local encoder (Swin-Tiny) produces a 7×7 grid of visual tokens [6]; (ii) a class-conditioned scorer ranks tokens and a grid-NMS [7] Top-K selector chooses diverse, non-redundant evidence; and (iii) a small global transformer (2 layer ViT-style) attends only to the selected K tokens plus [CLS] to predict the class. The same score map that gates selection is rendered as a heatmap; the K chosen coordinates are drawn as dots, aligning visualization with the actual evidence used by the classifier.

2.2 Local encoder and tokenization

Given an RGB input $x \in \mathbb{R}^{3\times224\times224}$, a tiny Swin transformer (window size 7) outputs the final feature map

$$F \in \mathbb{R}^{C\times H\times W}, \qquad C{=}768,\ H{=}W{=}7. \tag{1}$$

We flatten spatially and linearly project channels to D=384 to obtain token embeddings

$$T = \left[t_{ij}\right] \in \mathbb{R}^{N\times D}, \quad N{=}H\cdot W{=}49, \quad t_{ij} \;=\; \mathrm{Proj}(F_{:,i,j})\,. \tag{2}$$

A learnable class token $c \in \mathbb{R}^D$ (denoted [CLS]) is used in the global stage.

2.3 Class-conditioned scoring

Each token t_{ij} is assigned a class-conditioned score via cosine similarity between a linearly transformed token and the [CLS] query:

$$s_{ij} \;=\; \left\langle \frac{W\,t_{ij}}{\|W\,t_{ij}\|_2},\ \frac{c}{\|c\|_2} \right\rangle \in [-1,1], \qquad W \in \mathbb{R}^{D\times D}. \tag{3}$$

Reshaping the scores gives a heatmap $S \in \mathbb{R}^{H\times W}$ over the 7×7 grid. This heatmap is later visualized directly.

2.4 Diverse Top-K selection via grid-NMS

High scores can cluster spatially, leading to redundant evidence. To enforce spatial diversity, we apply greedy NMS on the grid with radius r=1: at each step we choose the current maximum of S and suppress its 3×3 neighborhood, repeating until K locations are retained. On a 7×7 lattice with r=1, the theoretical cap on unique non-overlapping picks is $\lceil H/2\rceil\cdot\lceil W/2\rceil = 16$, hence we set K=16 by default. Let $P = \{(i_\ell, j_\ell)\}_{\ell=1}^{K}$ be the selected coordinates and

$$Z = \left[z_\ell\right]_{\ell=1}^{K}, \qquad z_\ell \;=\; t_{i_\ell j_\ell} \in \mathbb{R}^D, \tag{4}$$

the corresponding evidence tokens forwarded to the classifier. Each selected token is augmented with a learned 2D positional embedding derived from its grid coordinates.

Method	Attended tokens	Global blocks	Acc. (%)	Macro-F1
X-TASAR	$K{+}1{=}17$	2	**98.05**	**0.979**
ViT-Tiny (full global)	$196{+}1$	12	*97.89*	*0.972*

Table 1: Results on the RGB-ArSLR alphabet test set when $K = 16$

2.5 Global transformer over `[CLS]`+K tokens

We build the sequence

$$X \;=\; \big[\, c,\; \tilde{z}_1, \ldots, \tilde{z}_K \,\big] \in \mathbb{R}^{(K+1)\times D}, \tag{5}$$

and process it with a small ViT-style transformer (two encoder blocks, six heads). Each encoder applies multi-head self-attention (MHSA), residual connections, and a position-wise MLP:

$$X' \;=\; X + \mathrm{MHSA}(\mathrm{LN}(X)), \qquad Y \;=\; X' + \mathrm{MLP}(\mathrm{LN}(X')). \tag{6}$$

We read the final class token $h_{\texttt{cls}} = Y_{1,:} \in \mathbb{R}^D$ and obtain the prediction

$$\hat{y} \;=\; \mathrm{softmax}(W_{\mathrm{cls}}\, h_{\texttt{cls}}), \qquad W_{\mathrm{cls}} \in \mathbb{R}^{C_y \times D}, \tag{7}$$

where C_y is the number of classes. Training uses cross-entropy on $\hat{y}$. For interpretability, we visualize S as a colored heatmap and overlay the K coordinates in P as dots, which are the exact patches consumed by the classifier.

3 Experiments

3.1 Experimental Setup

We evaluate the performance on the RGB ArSLR dataset [2] with 32 classes and 54049 RGB images collected from more than 40 signers. Images are resized to $224{\times}224$ and normalized per-channel. We use a fixed *train/val/test* partition with class-stratified sampling; [1]. We report Top-1 accuracy and macro-F1 on the test set.

Augmentations are random resized crop to 224, light color jitter, and random horizontal flip disabled (to preserve left/right handshape identity). We train with AdamW (weight decay 0.05) and cross-entropy loss. The best checkpoint is selected by validation macro-F1 and evaluated once on the test set. The baseline (ViT-Tiny) use the same optimizer, schedule, augmentations, epochs, and selection of the best checkpoint. Experiments were conducted in PyTorch with `timm` backbones on a multi-GPU Ubuntu server ($5\times$ NVIDIA RTX 3080, 24 GB VRAM each), with mixed-precision enabled. Exact dependencies are provided in the repository where We release the full codebase and the exact train/validation/test split CSVs used in this study https://github.com/brai-acslab/X-TASAR.

3.2 Baselines and Main Results

We compare X-TASAR to a ViT-Tiny baseline trained under the same protocol mentioned previously. The baseline is a 12-layer ViT that attends to all $14{\times}14{=}196$ patch tokens plus `[CLS]`. In contrast, our model scores the $7{\times}7$ grid, keeps a diverse Top-K set of patches ($K{=}16$) via grid-NMS, and runs a short 2-layer transformer only on $[\texttt{CLS}]{+}K$ tokens ($\approx$ 17 tokens total). As Table 1 shows, X-TASAR attains 98.05% accuracy, and 0.979 macro-F1, which is on par with, if not surpassing significantly, the full-global baseline while operating on fewer tokens.

3.3 Explainability

Beyond the numbers in Table 1, a central aim of X-TASAR is to make the model's internal evidence. Figure 2 presents randomly selected test images with our explainability overlays. The heatmap is the class-conditioned score map S over the $7{\times}7$ token grid, and the dots mark the Top-K ($K{=}16$) tokens chosen by grid-NMS with radius $r{=}1$. These dots correspond exactly to the tokens that are

[1] We release the exact CSVs (`train.csv`, `val.csv`, `test.csv`) used in this paper together with code at xxx

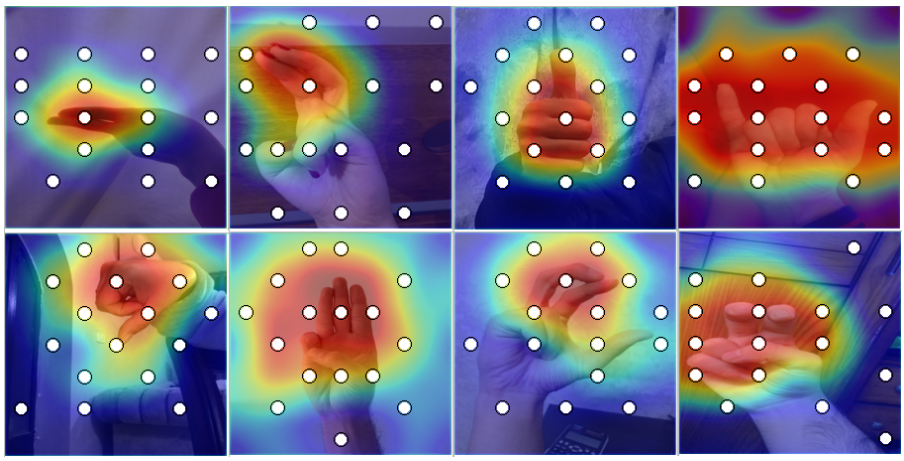

Figure 2: Heatmap with dots at the K selected patches (K=16, r=1). Warmer regions indicate higher class-conditioned scores; dots are the exact tokens passed to the global head.

subsequently passed, together with `[CLS]`, to the global transformer. This one-to-one linkage makes it easier to correlate the highlighted regions and marked locations that drive the classifier's decision.

In most cases the responses concentrate on the active hand and fingertip articulations, and the selected dots are well spread across complementary subregions, reflecting the enforced spatial diversity. Occasionally, some dots land on background or non-hand areas. Because selection operates on a coarse 7×7 grid with grid-NMS (r=1), very fine details (e.g., fingertips) that lie near cell boundaries have their evidence distributed over adjacent cells. Once one neighbor is selected, NMS suppresses the others, preventing multiple adjacent picks of the same fine structure. Combined with a fixed K value, this can yield a few residual selections on medium-score background cells. This behavior is expected from the proposed design and is visible in the overlays, which still depict exactly the tokens used by the classifier.

4 Limitations

The model has two main hyperparamers: the number of selected tokens K and the grid-NMS radius r. Smaller K risks missing discriminative parts, whereas larger K increases coverage at the cost of reintroducing background noise; similarly, a larger r enforces diversity but can thin out relevant or genuinely informative neighborhoods. In this study we fixed K=16 and r=1 based on validation, but data-driven or adaptive schemes that tune these values per sample remain a natural extension. A second limitation is the dimension of the feature map extracted. The 7×7 grid is computationally feasible but can be coarse for subtle finger articulations. Higher-resolution selection (e.g., from a 14×14 stage) or a multi-scale variant would likely sharpen localization and improve overlays, at an additional cost. Finally, the scope of evaluation is limited to RGB ArSLR alphabets dataset [2]. We do not yet evaluated the model on a continuous sign streams and other sign languages. This submission is intended as an early dissemination to elicit constructive feedback, towards the goal of expanding along the mentioned limitations.

5 Conclusion

We presented X-TASAR, an explainability-first transformer for isolated ArSLR recognition that scores a Swin-derived 7×7 grid, selects a diverse Top-K subset via grid-NMS, and classifies with a short 2-layer ViT-style head over $[\texttt{CLS}]+K$ tokens. The same score map drives both selection and visualization, yielding faithful heatmaps and dots, and the method attains strong performance (98.05% accuracy and 0.979 macro-F1) while greatly reducing the burden of global attention. In future, we will explore adaptive token selection that enhances discriminative points surrounding active hand regions, and broaden our evaluation under wider selection of SLR datasets to further benefit Arabic-speaking deaf and hard-of-hearing communities worldwide.

References

[1] Bashaer A Al Abdullah, Ghada A Amoudi, and Hanan S Alghamdi. Advancements in sign language recognition: A comprehensive review and future prospects. *IEEE Access*, 12:128871–128895, 2024. URL: https://doi.org/10.1109/ACCESS.2024.3457692.

[2] Muhammad Al-Barham, Adham Alsharkawi, Musa Al-Yaman, Mohammad Al-Fetyani, Ashraf Elnagar, Ahmad Abu SaAleek, and Mohammad Al-Odat. RGB Arabic Alphabets Sign Language Dataset, 2023. URL: https://arxiv.org/abs/2301.11932.

[3] Muhammad Al-Qurishi, Thariq Khalid, and Riad Souissi. Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. *IEEE Access*, 9:126917–126951, 2021. URL: https://doi.org/10.1109/ACCESS.2021.3110912.

[4] Ming Jin Cheok, Zaid Omar, and Mohamed Hisham Jaward. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, 10(1):131–153, 2019. URL: https://doi.org/10.1007/s13042-017-0705-5.

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL: https://arxiv.org/abs/2010.11929, arXiv:2010.11929.

[6] Ze Liu, Yutong Lin, Yue Cao, et al. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proc. ICCV*, pages 10012–10022, 2021. URL: https://arxiv.org/abs/2103.14030.

[7] A. Neubeck and L. Van Gool. Efficient Non-Maximum Suppression. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 3, pages 850–855, 2006. URL: https://doi.org/10.1109/ICPR.2006.479.

[8] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794, 2021. URL: https://doi.org/10.1016/j.eswa.2020.113794.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in abstract is reflected in Introduction and Experiments section.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: There is a separate Limitations section provided in the paper

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: Our approach focuses on enhancing the practical results and no theoretical aspects are presented. Hence, full set of assumptions and complete proof are not applicable

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 3 (Experiments) provides necessary information to reproduce the main experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper provides reference to the data used in the experiment. Additionally, access to the code base is provided with sufficient instructions reproduce the results

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specifies training, testing and validation splits and provides exact files used in training and validation as CSV files in the released code base.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Section 3 presents experiments that support the main claims of the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Section 3.1 details information on the compute resources.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The dataset used in the papaer was already anonymized prior to acquiring it. Hence, there are no personally identifiable information in the dataset.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The paper does not have direct societal impacts as it is designed for sign language recognition tasks.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The release code has no risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The data used in the study is credited with appropriate citation ([2])

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code used in the study has been released through GitHub repository and is mentioned in Section 3.1

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The study involves no crowdsourcing experiments and research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The LLM was used for editing the initial content written by authors.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.