ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

№2
2026

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

2 (358)

APRIL – JUNE 2026

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=6506682964, https://www.webofscience.com/wos/author/record/1423665

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=56153126500, https://www.webofscience.com/wos/author/record/2428551

MAMYRBAEV Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=55967630400, https://www.webofscience.com/wos/author/record/1774027

BAIGUNCHEKOV Zhumadil Zhanabaevich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=6506823633, https://www.webofscience.com/wos/author/record/1923423

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), https://www.scopus.com/authid/detail.uri?authorId=7005121594, https://www.webofscience.com/wos/author/record/678586

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), https://www.scopus.com/authid/detail.uri?authorId=56249263000, https://www.webofscience.com/wos/author/record/1268523

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=8701101900, https://www.webofscience.com/wos/author/record/1436451

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=37461441200, https://www.webofscience.com/wos/author/record/1768515

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), https://www.scopus.com/authid/detail.uri?authorId=56036884700, https://www.webofscience.com/wos/author/record/747649

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=55453992600, https://www.webofscience.com/wos/author/record/3802041

USATOVA Olga Alexandrovna, PhD, Associate Professor, Chief Scientific Secretary of the Institute of Information and Computing Technologies of the National Academy of Sciences of the Republic of Kazakhstan (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=57204581062, https://www.webofscience.com/wos/author/record/JCO-3058-2023

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), https://www.scopus.com/authid/detail.uri?authorId=57191242124,

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), https://www.scopus.com/authid/detail.uri?authorId=7202799321, https://www.webofscience.com/wos/author/record/38481396

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), https://www.scopus.com/authid/detail.uri?authorId=7004159952, https://www.webofscience.com/wos/author/record/46249977

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), https://www.scopus.com/authid/detail.uri?authorId=7006315935, https://www.webofscience.com/wos/author/record/524462

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
Certificate **№ KZ77VPY00121154** on the re-registration of the periodical printed and online publication of the information agency, issued on **05.06.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan
Subject area*: information and communication technologies.*
Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*
Periodicity: *4 times a year.*
http://www.physico-mathematical.kz/index.php/en/

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

МҰТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=6506682964, https://www.webofscience.com/wos/author/record/1423665

РЕДАКЦИЯ АЛҚАСЫ:

ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=56153126500, https://www.webofscience.com/wos/author/record/2428551

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=55967630400, https://www.webofscience.com/wos/author/record/1774027

БАЙГУНЧЕКОВ Жұмаділ Жаңабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=6506823633, https://www.webofscience.com/wos/author/record/1923423

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), https://www.scopus.com/authid/detail.uri?authorId=7005121594, https://www.webofscience.com/wos/author/record/678586

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), https://www.scopus.com/authid/detail.uri?authorId=56249263000, https://www.webofscience.com/wos/author/record/1268523

КЕЙЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=8701101900, https://www.webofscience.com/wos/author/record/1436451

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=37461441200, https://www.webofscience.com/wos/author/record/1768515

ОТМАН Мохаммед, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), https://www.scopus.com/authid/detail.uri?authorId=56036884700, https://www.webofscience.com/wos/author/record/747649

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=55453992600, https://www.webofscience.com/wos/author/record/3802041

УСАТОВА Ольга Александровна, PhD, қауымдастырылған профессор, ҚР ҒЖБМ "Ақпараттық және есептеу технологиялары институтының" бас ғалым хатшысы (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=57204581062, https://www.webofscience.com/wos/author/record/JCO-3058-2023

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), https://www.scopus.com/authid/detail.uri?authorId=57191242124,

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), https://www.scopus.com/authid/detail.uri?authorId=7202799321, https://www.webofscience.com/wos/author/record/38481396

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), https://www.scopus.com/authid/detail.uri?authorId=7004159952, https://www.webofscience.com/wos/author/record/46249977

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), https://www.scopus.com/authid/detail.uri?authorId=7006315935, https://www.webofscience.com/wos/author/record/524462

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).
Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025** ж. берген **№ KZ77VPY00121154** Куәлік.
Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*
Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
http://www.physico-mathematical.kz/index.php/en/

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимкаир Мутанович, доктор технических наук, профессор, академик НАН РК, (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=6506682964, https://www.webofscience.com/wos/author/record/1423665

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=56153126500, https://www.webofscience.com/wos/author/record/2428551

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=55967630400, https://www.webofscience.com/wos/author/record/1774027

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сатпаева (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=6506823633, https://www.webofscience.com/wos/author/record/1923423

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), https://www.scopus.com/authid/detail.uri?authorId=7005121594, https://www.webofscience.com/wos/author/record/678586

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), https://www.scopus.com/authid/detail.uri?authorId=56249263000, https://www.webofscience.com/wos/author/record/1268523

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=8701101900, https://www.webofscience.com/wos/author/record/1436451

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=37461441200, https://www.webofscience.com/wos/author/record/1768515

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), https://www.scopus.com/authid/detail.uri?authorId=56036884700, https://www.webofscience.com/wos/author/record/747649

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=55453992600, https://www.webofscience.com/wos/author/record/3802041

УСАТОВА Ольга Александровна, PhD, ассоциированный профессор, Главный ученый секретарь «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=57204581062, https://www.webofscience.com/wos/author/record/JCO-3058-2023

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), https://www.scopus.com/authid/detail.uri?authorId=57191242124,

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), https://www.scopus.com/authid/detail.uri?authorId=7202799321, https://www.webofscience.com/wos/author/record/38481396

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), https://www.scopus.com/authid/detail.uri?authorId=7004159952, https://www.webofscience.com/wos/author/record/46249977

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), https://www.scopus.com/authid/detail.uri?authorId=7006315935, https://www.webofscience.com/wos/author/record/524462

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*
Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания **№ KZ77VPY00121154.** Дата выдачи **05.06.2025**
Тематическая направленность: *информационно-коммуникационные технологии.*
В настоящее время: *вошел в список журналов, рекомендованных КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*
Периодичность: *4 раза в год.*
http://www.physico-mathematical.kz/index.php/en/

© ТОО «Центрально-азиатский академический научный центр», 2026

CONTENTS

COMPUTER SCIENCE

МАЗМҰНЫ

КОМПЬЮТЕРЛІК ҒЫЛЫМДАР

СОДЕРЖАНИЕ

КОМПЬЮТЕРНЫЕ НАУКИ

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 2.
Namber 358 (2026). 44–66

https://doi.org/10.32014/2026.2518-1726.426

IRSTI 28.23.15
UDC 004.932.2

© Akbarov D.[1], Sembayev T.[1], Kydyrbekova A.[1], 2026.
Astana IT University, Astana, Kazakhstan.
E-mail: 242672@astanait.edu.kz

QUALITY-AWARE POSE–HAND KEYPOINT EXTRACTION PIPELINE FOR SKELETON-BASED SIGN LANGUAGE RECOGNITION

Akbarov Damir — Astana IT University, Astana, Kazakhstan,
E-mail: 242672@astanait.edu.kz, https://orcid.org/0009-0002-2651-4168;
Sembayev Talgat — PhD, Assistant Professor, Astana IT University, Astana, Kazakhstan,
E-mail: talgat.sembayev@astanait.edu.kz, https://orcid.org/0000-0003-2360-8767;
Kydyrbekova Aigerim — PhD, Senior Lecturer, Astana IT University, Astana, Kazakhstan,
E-mail: aigerim.kydyrbekova@astanait.edu.kz, https://orcid.org/0009-0001-2384-6266.

Abstract. Skeleton-based sign language recognition (SLR) uses pose and hand keypoints as a compact, background-invariant representation, but in real signing videos the keypoint stream produced by off-the-shelf estimators is often unreliable. Hand trajectories frequently contain missing detections, left/right identity swaps, duplicate outputs, temporal jitter, and occlusion-related drift. These artifacts are not merely preprocessing noise: they directly degrade the quality of downstream SLR training. This paper presents a quality-aware pose–hand keypoint extraction pipeline that treats keypoint generation as an auditable, reliability-aware stage rather than a hidden preprocessing step. The proposed method combines: (i) pose-guided left/right hand assignment with spatial gating, (ii) pose-conditioned second-pass region-of-interest (ROI) refinement for low-confidence or missing hands, (iii) lightweight short-gap temporal bridging (tracking/hold), (iv) explicit occlusion-consistent labeling, (v) plausibility checks to reject physically implausible configurations, and (vi) anchor-preserving postprocessing for short-gap filling and jitter reduction. Beyond cleaned keypoints, the pipeline exports per-hand reliability semantics (provenance and state) and a dataset-ready quality report based on proxy metrics, including coverage, confidence, recovery fractions, intervention rates, and temporal stability indicators. A fixed composite quality score is also defined to support reproducible ranking and filtering without manually annotated hand landmarks. Dataset-scale evaluation on Slovo (n = 20,400 clips) shows that the proposed full configuration substantially improves extraction reliability over a pass1-only baseline: the composite quality score increases from 0.684 to 0.794, any-hand coverage from 0.889 to 0.969, and both-hands coverage from 0.454 to 0.646, while MissingRate decreases from 0.329 to 0.193. Recovery analysis further shows that most robustness gains come from ROI-based second-pass recovery, with tracking acting mainly as a sparse fallback mechanism. The pipeline is designed for reproducible keypoint export, dataset curation, and reliability-aware SLR training.

Keywords: skeleton-based sign language recognition; hand keypoints; pose-guided recovery; temporal consistency; occlusion robustness; quality metrics; MediaPipe

For citations: Akbarov D., Sembayev T. Кыдырбекова А. Quality-aware pose–hand keypoint extraction pipeline for skeleton-based sign language recognition. Academic Scientific Journal of Computer Science, 2026. — No.2. – P. 44–66. DOI: https://doi.org/10.32014/2026.2518-1726.426

© Акбаров Д.Р.[1], Сембаев Т.М.[1], Кыдырбекова А.Б.[1], 2026.
Astana IT University, Астана, Казақстан.
E-mail: 242672@astanait.edu.kz

ЫМ ТІЛІН ТАНУҒА АРНАЛҒАН ДЕНЕ ҚАЛПЫ МЕН ҚОЛДЫҢ НЕГІЗГІ НҮКТЕЛЕРІН САПАНЫ БАҚЫЛАУМЕН АНЫҚТАУ ӘДІСІ

Акбаров Дамир — Astana IT University, Астана, Қазақстан,
E-mail: 242672@astanait.edu.kz, https://orcid.org/0009-0002-2651-4168;
Сембаев Талгат — PhD, ассистент-профессор, Astana IT University, Астана, Казақстан,
E-mail: talgat.sembayev@astanait.edu.kz, https://orcid.org/0000-0003-2360-8767.
Кыдырбекова Айгерим — PhD, Сеньор-лектор, Astana IT University, Астана, Казахстан,
E-mail: aigerim.kydyrbekova@astanait.edu.kz, https://orcid.org/0009-0001-2384-6266.

Аннотация. Қаңқалық деректерге негізделген ым тілін тану міндеттерінде кіріс ұсынылымы әдетте дене қалпы мен қолдың негізгі нүктелерінің уақыттық тізбегі түрінде беріледі, бұл сипаттаманы ықшамдап, фонға тәуелділікті азайтады. Алайда нақты бейнежазбаларды өңдеу кезінде негізгі нүктелерді бағалаудың стандартты құралдары беретін нәтижелерді тікелей пайдалану жеткілікті сенімді емес. Қол траекторияларында детекциялардың түсіп қалуы, сол және оң қол сәйкестігінің ауысуы, бағалаулардың қайталануы, координаттардың уақыт бойынша тұрақсыздығы және окклюзияларға байланысты ығысулар жиі кездеседі. Мұндай бұрмаланулар кейінгі тану модельдерін оқыту сапасына тікелей әсер етеді. Жұмыста дене қалпы мен қолдың негізгі нүктелерін сапаны бақылаумен қалыптастыру әдісі ұсынылады; онда координаттарды алу кезеңі сенімділіктің айқын аннотациясы бар, дербес тексерілетін және қайта жаңғыртылатын саты ретінде қарастырылады. Әдіс дене қалпына сүйеніп сол және оң қолды тағайындау мен кандидаттарды кеңістіктік сүзгілеуді, сенімділігі төмен не деректері жоқ кадрлар үшін қызығушылық аймағында (ROI) қайта нақтылауды, қысқа үзілістерді жабуға арналған жеңілдетілген уақытша ілестіріп қадағалау мен бағалауды ұстап тұруды, окклюзия кезіндегі бірізді таңбалауды, қисынға сәйкестік тексерістерін және тірек кадрларды сақтай отырып кейінгі өңдеуді біріктіреді. Тазартылған траекториялардан бөлек, әдіс әр қол үшін сенімділік семантикасын (бағалау көзі мен күйі), жанама метрикаларға негізделген сапа есебін және эталондық таңбалануы жоқ жағдайда деректерді ранжирлеу мен сүзгілеуге арналған қайта жаңғыртылатын құрама сапа көрсеткішін қалыптастырады. Slovo деректер жиынында (n = 20 400) толық конфигурация құрама сапа көрсеткішін 0,684-тен 0,794-ке дейін арттырып, MissingRate көрсеткішін 0,329-дан 0,193-ке дейін төмендетеді.

Түйін сөздер: қаңқалық деректерге негізделген ым тілін тану; қолдың негізгі нүктелері; негізгі нүктелер сапасын бақылау; дене қалпына сүйенген қалпына келтіру; уақытша келісімділік; жанама сапа метрикалары; MediaPipe

© Акбаров Д.Р.[1], Сембаев Т.М.[1], Кыдырбекова А.Б.[1], 2026.
Astana IT University, Астана, Казахстан.
E-mail: 242672@astanait.edu.kz

МЕТОД ПОЛУЧЕНИЯ КЛЮЧЕВЫХ ТОЧЕК ПОЗЫ И КИСТЕЙ С КОНТРОЛЕМ КАЧЕСТВА ДЛЯ РАСПОЗНАВАНИЯ ЖЕСТОВОГО ЯЗЫКА

Акбаров Дамир — Astana IT University, Астана, Казахстан,
E-mail: 242672@astanait.edu.kz, https://orcid.org/0009-0002-2651-4168;
Сембаев Талгат — PhD, ассистент-профессор, Astana IT University, Астана, Казахстан,
E-mail: talgat.sembayev@astanait.edu.kz, https://orcid.org/0000-0003-2360-8767;
Кыдырбекова Айгерим — PhD, Сеньор-лектор, Astana IT University, Астана, Казақстан,
E-mail: aigerim.kydyrbekova@astanait.edu.kz, https://orcid.org/0009-0001-2384-6266.

Аннотация. *Актуальность.* В задачах распознавания жестового языка по скелетным данным входное представление обычно задается временной последовательностью ключевых точек позы и кистей, что обеспечивает компактность описания и снижает зависимость от фона. Однако при обработке реальных видеозаписей прямое использование выходных данных стандартных средств оценки ключевых точек оказывается недостаточно надежным. Для траекторий кистей характерны пропуски детекций, перестановки левой и правой руки, дублирование оценок, временная нестабильность координат и смещения, обусловленные окклюзиями. Такие искажения нельзя рассматривать как несущественный шум предварительной обработки, поскольку они напрямую влияют на качество последующего обучения моделей распознавания. *Цель.* Разработать метод формирования ключевых точек позы и кистей с контролем качества для повышения надежности входных скелетных данных при распознавании жестового языка. *Методы.* В работе предложен метод формирования ключевых точек позы и кистей с контролем качества, в котором этап получения координат рассматривается как самостоятельная проверяемая и воспроизводимая стадия с явной аннотацией надежности. Метод объединяет назначение левой и правой кисти, пространственную фильтрацию кандидатов с опорой на позу, повторное уточнение в области интереса для кадров с низкой уверенностью или пропусками, краткосрочное временное сопровождение и удержание оценки. Также предусмотрены согласованная маркировка состояний при окклюзиях, проверки правдоподобия и постобработка для заполнения коротких разрывов и снижения нестабильности траекторий. *Результаты и выводы.* Дополнительно формируются семантика надежности для каждой кисти и отчет о качестве на основе косвенных метрик, включающих покрытие, уверенность, долю восстановления, частоту корректирующих вмешательств и показатели временной нестабильности. Также предложен воспроизводимый составной показатель для ранжирования и фильтрации данных при подготовке обучающих выборок без эталонной разметки ключевых точек. На датасете Slovo (n = 20 400) метод в полной конфигурации повышает составной показатель качества с 0,684 до 0,794 и снижает MissingRate с 0,329 до 0,193 по сравнению с базовым вариантом первого прохода обработки. Полученные результаты подтверждают эффективность предложенного подхода для повышения качества скелетных данных и подготовки обучающих выборок в задачах распознавания жестового языка.

Ключевые слова: распознавание жестового языка, скелетные данные, ключевые точки кистей, контроль качества ключевых точек, восстановление с опорой на позу, временная согласованность, косвенные метрики качества, MediaPipe.

Introduction. Skeleton-based sign language recognition (SLR) models signing as multivariate time series of body and hand landmarks. Compared with RGB, landmark sequences offer a compact, structured signal that is less sensitive to background, clothing, and identity, and they enable efficient temporal modeling focused on kinematics rather than texture. In practice, however, the effectiveness of skeleton-based SLR is bounded by the reliability of extracted keypoints, with hands being the dominant failure source due to fast motion, frequent self-occlusions, and limited image evidence for fingers.

A common workflow is to export per-frame detections from an off-the-shelf pose/hand estimator and treat the resulting keypoint stream as a preprocessing step. This direct extraction is systematically brittle under realistic signing conditions: hands may be missing or low-confidence, left/right identity can swap when hands cross, duplicate detections may appear, short detector dropouts fragment trajectories, and frame-to-frame jitter injects high-frequency noise. These artifacts are not benign. When SLR models are trained on corrupted keypoint sequences, they can overfit to systematic extraction artifacts, leading to degraded robustness and poor cross-domain generalization.

We treat keypoint extraction as a quality-critical stage and introduce a modular pipeline that improves hand landmark reliability without requiring ground-truth landmarks. The pipeline integrates pose-guided constraints with temporal consistency: (i) an assignment module that associates detected hands to the signer's left/right identity using body pose cues and continuity, (ii) a second-pass recovery that revisits low-quality segments to recover plausible trajectories under occlusion and detector failures, (iii) lightweight tracking/hold logic to bridge short gaps while limiting drift, (iv) sanity checks to reject physically implausible configurations and inconsistent motion, and (v) postprocessing to reduce jitter while preserving sign-relevant dynamics.

Equally important, we argue that reliable extraction must be accompanied by measurement. Without explicit quality metrics, dataset-scale extraction is effectively unobservable: practitioners cannot quantify how often hands are missing, how frequently swaps occur, or which videos are unusable. We therefore provide a quality report based on proxy metrics (coverage, confidence, continuity, anomaly rates, and recovery impact) to support filtering, curriculum design, and systematic debugging.

The paper makes the following contributions:

- a pose-guided left/right hand assignment procedure that reduces swaps and duplicates;
- a second-pass recovery strategy for occlusions and detector dropouts;
- a temporal tracking/hold module and sanity validation to enforce consistency;
- postprocessing tailored to suppress jitter without erasing motion cues;
- a quality-report protocol with proxy metrics for dataset-level monitoring and data selection.

Section 2 reviews related work on skeleton-based SLR and landmark reliability. Section 3 formulates the problem and defines the output representation (landmarks, reliability signals, and export format). Section 4 describes the proposed modular extraction pipeline (assignment, second-pass recovery, tracking/hold, sanity checks, and postprocessing). Section 5 presents the quality-report protocol. Section 6 presents the dataset-scale evaluation and results. Section 7 discusses practical implications and limitations, and Section 8 concludes the paper.

Literature review. Research most relevant to quality-aware pose–hand keypoint extraction for skeleton-based SLR spans four areas: (1) keypoint-based SLR architectures, (2) pose/hand estimation and holistic extraction pipelines, (3) robustness to corrupted keypoint streams and temporal consistency, and (4) dataset and protocol effects. Here, *extraction quality* refers to measurable, extraction-time signals—ROI validity/overlap, per-keypoint confidence (e.g., visibility/presence), missingness, temporal jitter, and tracking state—that determine whether landmark trajectories are suitable for downstream learning.

Skeleton- and Keypoint-Based Sign Language Recognition

Skeleton-based SLR typically models signing as a spatio-temporal graph of body and hand landmarks. ST-GCN is a widely used baseline, applying graph convolutions over kinematic edges within frames and temporal edges across frames (Yan et al., 2018). Subsequent work emphasizes that graph design is task-dependent: SAM-SLR reports that focusing on signing-relevant joints (hands and upper body) can be preferable to dense whole-body graphs when additional landmarks contribute non-informative structure (Jiang et al., 2021). At the word level, models such as GCN-BERT separate per-frame spatial encoding from temporal modeling via transformer attention, but remain sensitive to viewpoint changes and out-of-plane motion under 2D keypoints (Tunga et al., 2021). Temporal summarization approaches further show degradation under speed variation and temporal misalignment of hold phases, indicating that robust SLR benefits from stable trajectories and explicit treatment of missing or noisy keypoints (Kindiroglu et al., 2019). Overall, this literature largely assumes that the keypoint stream is sufficiently reliable and rarely standardizes extraction-time quality signals that would enable comparable training and evaluation across pipelines.

Pose/Hand Estimation and Holistic Pipelines

Practical extraction pipelines commonly rely on cascaded detection–tracking with ROI normalization. OpenPose introduced bottom-up multi-person pose estimation using Part Affinity Fields, but scale variation and crowding can propagate association errors and noise downstream (Cao et al., 2017). Production systems therefore often adopt detector–tracker designs implemented in pipeline frameworks such as MediaPipe, enabling large-scale processing with intermittent detection and lightweight tracking between detections (Lugaresi et al., 2019), while models such as BlazePose additionally output per-keypoint visibility/reliability under partial views and occlusions (Bazarevsky et al., 2020).

For hands, MediaPipe Hands detects palms and regresses 21 landmarks from a cropped ROI, re-running detection when a learned presence score indicates tracking failure; ROI drift and occlusion remain key sources of landmark degradation (Zhang et al., 2020). Holistic variants incorporate refinement via high-resolution hand re-cropping guided by pose cues (Grishchenko et al., 2022), and targeted analysis of MediaPipe Holistic shows that default ROI heuristics can fail under non-ideal orientations while learned ROI estimators improve overlap and worst-case behavior (Moryossef, 2024). These works motivate ROI-aware extraction, but do not establish a dataset-scale layer that consistently standardizes ROI control, temporal constraints, sanity validation, and unified quality reporting across large runs.

Robustness to Missing/Noisy Keypoints and Temporal Consistency

Robust skeleton modeling addresses occlusion, missingness, and jitter through reliability-aware gating and denoising objectives (Liu et al., 2016; Song et al., 2021; Paoletti et al., 2023). In SLR, tracking error is associated with recognition degradation, and using tracked hand regions can reduce word error rate (WER) compared with coarse full-frame cues (Koller et al., 2015). In real-time settings, missing keypoints are often handled with heuristics rather than explicit missing-data modeling, which can introduce discontinuities and bias temporal features (Moryossef et al., 2020).

However, most robustness approaches focus on recognition-time mitigation given a fixed skeleton stream. They typically do not enforce extraction-time temporal consistency, detect/flag trajectory failures, or produce standardized QC outputs that can be used for dataset curation and reproducible training.

Datasets, Protocols, and Data Quality

Large-scale datasets amplify sensitivity to preprocessing and extraction choices. WLASL provides a major word-level benchmark with metadata and pose-based baselines, but landmarks are commonly extracted via off-the-shelf pipelines without standardized reliability reporting, making it difficult to separate extraction effects from modeling effects (Li et al., 2020). How2Sign provides a large-scale continuous American Sign Language corpus with parallel modalities and automatically extracted keypoints (including face), and reports that extraction settings such as resolution influence confidence, particularly for hands, motivating explicit auditing of landmark quality (Duarte et al., 2021; Cao et al., 2017). Slovo provides large-scale Russian Sign Language data but does not include pose/hand keypoints or quality labels, making robust external extraction important for skeleton-based training and evaluation (Kapitanov et al., 2023). Consequently, results can vary substantially with ROI/cropping policies, tracking behavior, and missingness handling, and remain difficult to compare without dataset-level QC reporting.

Prior work offers strong keypoint-based recognizers, practical extraction components, and recognition-time robustness strategies, but lacks a unified, dataset-scale pose–hand extraction layer that standardizes ROI quality control, enforces temporal consistency with sanity validation, and exports consistent per-frame and dataset-level quality reports for reproducible training and evaluation.

Problem Formulation and Output Schema. We process each sign-language clip into a time-aligned pose–hand stream for the primary signer. The target output is directly consumable by skeleton-based SLR models and is designed for unconstrained footage, where hand tracks are prone to missed detections, left–right identity swaps, drift, and self-occlusion. To make the output usable under such conditions, the extractor provides not only landmark coordinates but also per-hand reliability annotations that enable downstream masking, weighting, and conditioning.

The extractor produces an ordered sequence of processed timesteps, optionally temporally subsampled. Each timestep contains an upper-body pose estimate when available, left and right hand keypoints when available, and per-hand metadata consisting of a confidence score and two categorical reliability labels.

Skeleton Stream Specification

For each processed timestep, the stream contains one pose component and two hand components. The pose corresponds to a fixed upper-body subset of joints, with the joint set and ordering fixed per run. Each hand is represented using a canonical 21-landmark convention with a fixed landmark ordering. A scalar confidence score is provided per hand.

When any component is unavailable, it is explicitly marked as missing via a validity flag and handled consistently by the pipeline's reliability logic. The pose subset additionally provides global motion context and wrist priors that help stabilize hand localization under detector dropouts and partial occlusions.

The extractor operates in a single coordinate mode per run to ensure consistency and reproducibility at dataset scale. In image mode, coordinates are stored as normalized 2D values relative to frame width and height; when supported, an additional depth-like channel may be included. In world mode, 3D coordinates are stored when available. For efficiency, pose may be computed less frequently than hands; intermediate timesteps may reuse the nearest available pose estimate, and such reuse is recorded explicitly.

Quality-Aware Pose–Hand Extraction Pipeline. Building on the output representation defined in Section 3, we describe a quality-aware pipeline that constructs the pose–hand stream and assigns per-hand provenance and reliability labels at every timestep. The method is organized as a sequence of quality-control stages that progressively refine hand estimates and suppress common failure modes. Concretely, we first obtain per-frame pose and hand hypotheses from a primary detector (pass1). Pose-derived wrist priors are used to resolve left/right identity and reject spatially implausible hand outputs. When pass1 is weak or missing, a pose-conditioned second-pass ROI refinement performs localized re-detection to recover difficult frames. Residual short failures are bridged via temporal robustness mechanisms (tracked/hold) and occlusion-consistent handling, followed by lightweight plausibility checks to suppress catastrophic artifacts. Finally, an optional anchor-based postprocessing step fills short gaps and smooths trajectories while preserving high-confidence anchor frames (Figure 1).

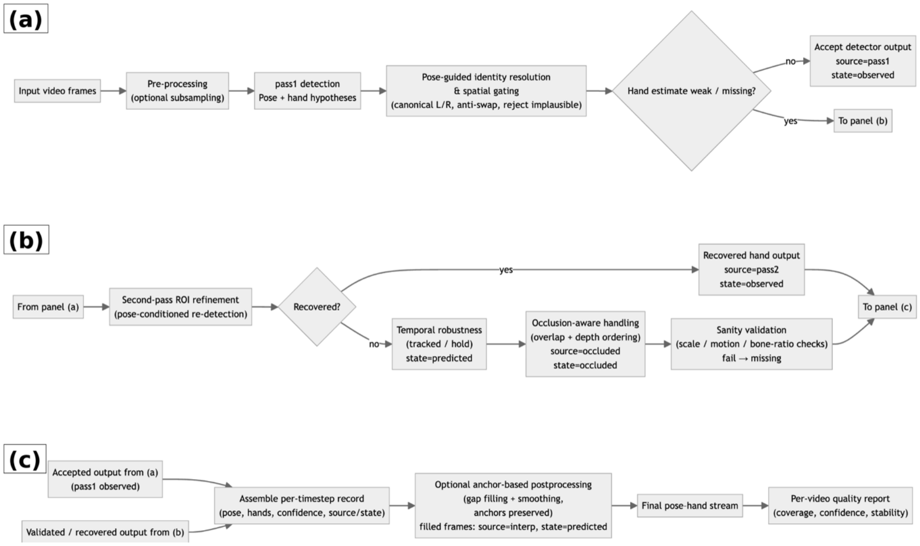

Figure 1. Overview of the quality-aware pose–hand extraction pipeline: (a) pass1 detection with pose-guided identity resolution and spatial gating; (b) ROI-based recovery (pass2), temporal robustness (tracked/hold), occlusion-consistent handling, and sanity validation; (c) stream assembly, optional anchor-preserving postprocessing (interp), and per-video quality reporting.

Reliability-Aware Stream Construction

Rather than treating the per-hand confidence $c_t^h \in [0,1]$ as a passive diagnostic, we use it as a control signal that selects between extraction modes. High-confidence detections are accepted as directly observed evidence, whereas weak or missing detections trigger recovery. Specifically, if a hand hypothesis is absent or falls below a predefined trigger threshold, the pipeline performs a localized second-pass ROI re-detection centered on a pose-conditioned hand prior. If recovery fails, the pipeline falls back to short-range temporal propagation (tracked/hold) or marks the hand as missing. Confidence is derived either from the detector's handedness-related score (used here as a proxy confidence, default) or from a binary presence indicator when score calibration is unreliable; in both cases, downstream semantics are carried by the reliability labels (u_t^h, q_t^h).

Reliability is operationalized by separating provenance (source) from interpretation (state). The source label u_t^h identifies which mechanism produced the final hand estimate at timestep t, while the state label q_t^h provides a compact reliability regime that is directly actionable downstream. Detector-derived outputs (pass1, pass2) are labeled as observed; temporally propagated outputs (tracked, hold) are labeled as predicted; and overlap-driven, occlusion-consistent outputs are labeled as occluded. When no estimate is available, the state is missing and provenance is undefined (we set $u_t^h = \emptyset$). Postprocessing may additionally synthesize short-gap estimates via anchor-based interpolation; such frames are labeled $u_t^h = \mathrm{interp}$ and interpreted as $q_t^h = \mathrm{predicted}$, preserving consistent downstream handling. This separation enables SLR models to mask missing segments, down-weight inferred intervals, or explicitly condition on reliability regimes rather than assuming uniform keypoint quality. Table 1 summarizes the source–state mapping and recommended downstream usage.

Finally, we define anchors as high-confidence, sanity-validated observed frames that serve as stable reference points for subsequent stabilization. Anchors are used (i) as reference configurations for plausibility checks (e.g., scale and bone-ratio consistency) and (ii) as fixed points during optional postprocessing, where short-gap filling and smoothing are applied in an anchor-preserving manner to avoid distorting reliable observations. For auditability, the pipeline records rejection diagnostics when candidates fail plausibility constraints, enabling dataset-scale inspection of typical failure modes. Anchors also delimit anchor-protected gap filling in postprocessing: any filled frames remain explicitly marked as inferred ($u_t^h = \text{interp}, q_t^h = \text{predicted}$) rather than being merged into observed evidence.

Table 1 — Reliability semantics used in the pipeline and recommended downstream usage.

Source (u_t^h)	**State (q_t^h)**	**Origin (how produced)**	**Suggested usage in SLR**
pass1	observed	primary detector output	full weight; eligible anchor
pass2	observed	pose-conditioned ROI re-detection	full/near-full weight; eligible anchor
tracked	predicted	temporal propagation via tracking	reduced weight or separate token; do not treat as ground truth
hold	predicted	short-gap reuse of last valid estimate	reduced weight; allow only short spans
interp	predicted	anchor-based short-gap interpolation (postprocess)	reduced weight; treat as inferred; allow only short spans
occluded	occluded	occlusion-consistent estimate under overlap ambiguity	optionally down-weight; condition model on occlusion regime
$\emptyset$	missing	no estimate available	hard mask / ignore for that hand

Pose-Guided Hand Identity Resolution and Spatial Gating

Off-the-shelf hand detectors frequently exhibit left/right identity flips and occasional spurious or duplicated hypotheses, especially under rapid motion, partial self-occlusion, and boundary truncation. To enforce a stable canonical left/right ordering and suppress physically implausible outputs before downstream recovery, we use pose wrists as geometric priors and apply pose-guided identity resolution with spatial gating (Figure 2).

Let the left/right pose wrist references at time t be $w_p^L(t)$ and $w_p^R(t)$, and let $w(H)$ denote the wrist landmark of a detected hand hypothesis H. We compare hypotheses using a scale-normalized wrist proximity,

$$d_h(H,t) = \frac{\| w(H) - w_p^h(t) \|_2}{\max\left(s(H), \epsilon\right)}, h \in \{L, R\}. \tag{1}$$

where $s(H)$ is an intrinsic hand-scale estimate (from landmark spread) to reduce sensitivity to subject scale and camera zoom. When two hypotheses are available, we evaluate both the no-swap and swapped assignments and select the configuration with lower total proximity cost (Figure 2a). To avoid unstable decisions when pose is unreliable, pose-guided assignment is applied only when wrist priors are valid; otherwise, the detector-provided handedness is retained.

When only a single hypothesis is present, we assign it to the side whose pose wrist yields the smaller normalized distance; if it is markedly closer to the opposite wrist than to its current label, we reassign it (Figure 2b). Finally, we apply spatial gating by rejecting hypotheses that are too far from their assigned wrist in normalized units (Figure 2c), and we suppress duplicates by retaining the higher-confidence hypothesis when two outputs are near-identical under IoU/landmark similarity (Figure 2d).

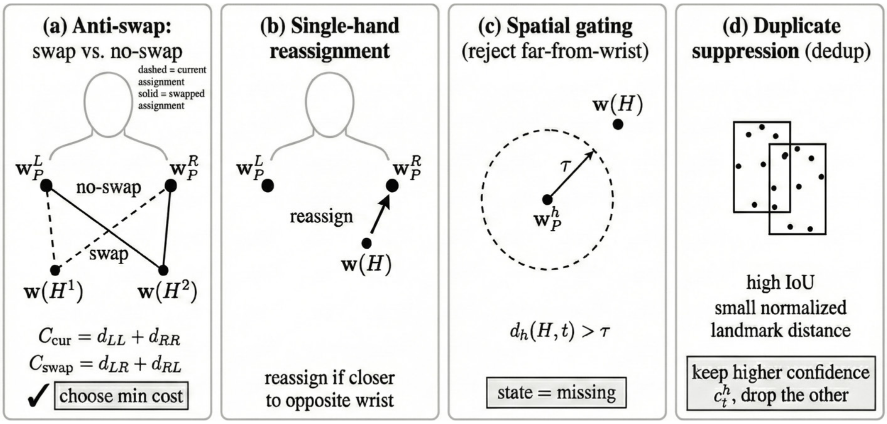

Figure 2 — Pose-guided hand identity resolution and spatial filtering: anti-swap assignment, single-hand reassignment, distance-based gating ($d_h(H,t) > \tau$; missing), and duplicate suppression (IoU/landmark similarity), keeping the higher-confidence hypothesis c_t^h.

Second-Pass Pose-Conditioned ROI Refinement

Pass1 provides the primary hand hypotheses, but in unconstrained signing it remains the dominant failure point due to motion blur, truncation, and self-occlusion. To improve recall while preserving the reliability semantics of Section 4.1, we introduce a second-pass refinement that performs localized re-detection in a pose-conditioned ROI. The second pass is invoked only when the pass1 output for side $h \in \{L, R\}$ is missing or falls below a trigger threshold.

For each side h, the ROI is centered at an expected hand location derived from the pose wrist prior $w_P^h(t)$. If the wrist prior is unavailable or unreliable, the center falls back to the most recent valid hand location (and, if necessary, a neutral default), ensuring stable behavior under intermittent pose failures (Figure 3). Because the wrist prior can be imprecise, we search a small ladder of increasing ROI scales and, at each scale, probe a compact set of local center offsets (jitter). The search proceeds from smaller to larger ROIs and terminates at the first scale that yields a valid candidate, keeping computation bounded while covering typical prior error (Figure 3).

Figure 3. Second-pass pose-conditioned ROI refinement flow. When pass1 is missing or weak, the method centers an ROI at the expected hand location (pose wrist with fallbacks), searches a small multi-scale and jitter neighborhood, and selects the best candidate via a confidence–center-consistency score, yielding $u_t^h = \mathrm{pass2}$ and $q_t^h = \mathrm{observed}$ upon success.

Each candidate hypothesis H provides a detector confidence $c(H)$ and handedness label. We select the best candidate using a confidence–center-consistency score,

$$S(H) = c(H) - \lambda\delta(H), \quad \delta(H) = \frac{\| w(H) - \mu \|_2}{s_{roi}}, \tag{2}$$

where μ is the ROI center and s_{roi} is the ROI size used for normalization. Candidates with excessive deviation are rejected; handedness preference is soft to avoid systematic misses when the detector's handedness is uncertain. Upon success, the recovered estimate is labeled as pass2/observed; otherwise the pipeline defers to subsequent temporal handling or missing-state output.

Temporal Robustness

Detector outputs in unconstrained signing are often intermittent: hands may disappear briefly, reappear with jitter, or become ambiguous under overlap. To preserve temporal continuity over bounded gaps while keeping reliability explicit, we apply lightweight temporal mechanisms that populate the stream with clearly labeled predicted and occluded intervals when direct observations are weak or missing.

When a reliable hand estimate is available but subsequent frames exhibit short detector failures, we propagate the last valid hand using optical-flow correspondences with a constrained similarity transform to preserve hand shape, and apply a lightweight temporal filter to reduce high-frequency jitter. Frames produced by tracking are labeled as tracked/predicted. If tracking is unavailable or unstable and the dropout is very short, we use a conservative hold strategy that reuses the last exported hand configuration for a bounded number of frames, labeled as hold/predicted.

Under strong hand–hand overlap, detector outputs may flicker and left/right identity may become unstable. In this regime, we detect occlusion using high 2D overlap (intersection-over-union, IoU) together with depth-ordering cues when available, and emit an occlusion-consistent estimate that preserves continuity while explicitly exposing ambiguity; such frames are labeled as occluded/occluded. To prevent catastrophic artifacts from propagating through temporal recovery, we apply plausibility constraints (e.g., scale consistency relative to the last reliable anchor, wrist-jump bounds, and bone-ratio stability); failing candidates are rejected and recorded for auditability. Crucially, these mechanisms do not overwrite detector-derived observed estimates and only fill explicitly labeled predicted/occluded intervals.

Anchor-Based Postprocessing

While the preceding stages already produce an explicit reliability-annotated stream, brief dropouts and frame-to-frame jitter may still reduce temporal stability. We therefore apply an optional anchor-based postprocessing step that fills short gaps between reliable observations and smooths non-anchor segments while preserving anchors exactly.

For each hand h, we define anchors as high-confidence, sanity-valid *observed* estimates (i.e., $q_t^h = \mathrm{observed}$ from detector-derived sources $u_t^h \in \{\mathrm{pass1}, \mathrm{pass2}\}$). For any two consecutive anchors (t_0, t_1), we fill only short gaps whose length satisfies $(t_1 - t_0 - 1) \leq G_{\max}$. To preserve hand shape, each anchor configuration is represented by its wrist position p_t^h, a scalar hand scale s_t^h, and normalized landmark offsets

$$O_t^h = \frac{H_t^h - p_t^h}{\max(s_t^h, \epsilon)}, \tag{3}$$

Intermediate frames $t \in (t_0, t_1)$ are synthesized by linear interpolation of $\{p_t^h, s_t^h, O_t^h\}$ between the two anchors, followed by reconstruction $H_t^h = p_t^h + s_t^h O_t^h$. All filled frames are explicitly labeled as postprocessed interpolation, $u_t^h = \mathrm{interp}$ and $q_t^h = \mathrm{predicted}$, in accordance with the source–state mapping defined in Table 1.

After gap filling, we optionally smooth trajectories to reduce high-frequency jitter. Smoothing is applied only to non-anchor frames and not across long missing intervals; anchors remain unchanged (or are restored after smoothing) to preserve reliable detector evidence. Figure 4 illustrates the effect on a representative wrist trajectory segment.

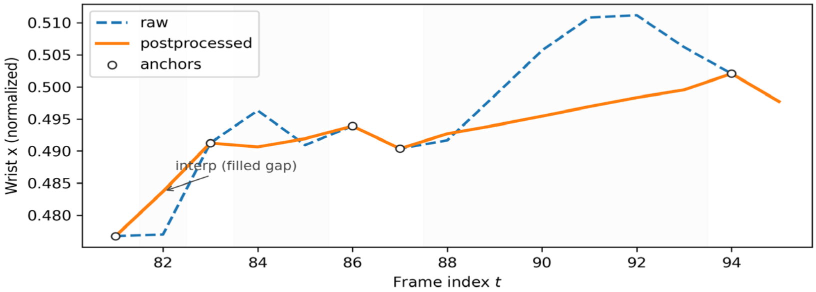

Figure 4. Anchor-based postprocessing example (zoomed segment): short gaps between anchors are filled by anchor-preserving interpolation, and optional smoothing reduces jitter on non-anchor frames while leaving anchors unchanged.

Quality Report: Definitions and Metrics. The extractor is designed to support large-scale sign-language dataset processing; however, the quality report is defined at the level of a single video and can be aggregated across a corpus when desired. Quality is characterized using observable proxies computed from the extracted stream, without requiring ground-truth hand landmarks. Let T′ denote the number of processed frames. We use an indicator function I(·) that equals 1 when its argument is true and 0 otherwise. We index hands by h ∈ {L, R}. A hand is considered present at frame t when the extractor outputs a non-empty landmark set; we denote this by $H_t^h \neq \emptyset$. Let $c_t^h \in [0, 1]$ denote the per-frame confidence of the extracted hand estimate, and let P_t denote the pose estimate at frame t. Finally, let u_t^h denote the source label of the hand estimate (Section 4.1). We report coverage and confidence statistics, a fixed composite score for ranking, and auxiliary diagnostic rates that explain how coverage is achieved and how stable the stream is over time.

Coverage

Per-hand coverage measures the fraction of frames in which each hand is present:

$$Cov_{\mathrm{L}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}(H_t^L \neq \emptyset) \tag{4}$$

$$Cov_{\mathrm{R}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}(H_t^R \neq \emptyset) \tag{5}$$

For compact reporting, we also define overall hand coverage as the mean of the per-hand coverages:

$$Cov_{any} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}(H_t^L \neq \emptyset \vee H_t^R \neq \emptyset) \tag{6}$$

Both-hands coverage measures the fraction of frames where both hands are simultaneously available, which is particularly important for two-handed signs:

$$Cov_{\mathrm{LR}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}(H_t^L \neq \emptyset \wedge H_t^R \neq \emptyset) \tag{7}$$

Pose coverage is defined analogously as the fraction of frames for which a pose estimate is available in the output stream. When pose is computed sparsely, the nearest available pose estimate is reused and still counted as available:

$$Cov_{\mathrm{P}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}(P_t \text{ exists}) \tag{8}$$

Composite quality score.

For dataset-scale ranking and filtering under a fixed extraction configuration, we define a per-video quality score Q as a weighted combination of both-hands coverage Cov_{LR} (Eq. 7), mean per-hand coverage $(Cov_L + Cov_R)/2$ (Eq. 4–5), pose coverage Cov_P (Eq. 8). For each hand $h \in \{L, R\}$, we define the median confidence over frames where the hand is present as:

$$\mathrm{MedConf}_h = \mathrm{median}\left(\{c_t^h : H_t^h \neq \emptyset\}\right) \tag{9}$$

In the quality score, we use the mean of the two instances of $\mathrm{MedConf}_h$ for $h \in \{L, R\}$, i.e., $(\mathrm{MedConf}_L + \mathrm{MedConf}_R)/2$. The weights are fixed to make the score reproducible and comparable across runs:

$$Q = 0.4 \cdot Cov_{\mathrm{LR}} + 0.2 \cdot \frac{(Cov_{\mathrm{L}} + Cov_{\mathrm{R}})}{2} + 0.2 \cdot Cov_{\mathrm{P}} + 0.2 \cdot \frac{(\mathrm{MedConf}_L + \mathrm{MedConf}_R)}{2} \tag{10}$$

This score is not intended as a universal measure of correctness; rather, it provides a reproducible proxy for how reliably a video yields usable pose–hand skeletons for skeleton-based SLR under a chosen extraction configuration.

Auxiliary diagnostic rates.

The composite score alone does not indicate whether coverage is obtained via direct detections or via recovery mechanisms, and it does not quantify identity corrections or temporal jitter. We therefore report additional diagnostic rates that are interpretable without ground truth and enable dataset-scale auditing and ablation studies.

Recovery fractions.

Using the source label u_t^h, we measure the fraction of frames where a hand estimate is produced by ROI-based second-pass refinement pass2, bridged by tracker-based temporal prediction tracked, or bridged by short-gap temporal propagation hold. The primary detector output pass1 is omitted as the default case:

$$\mathrm{RecFrac}^h{}_{\mathrm{pass2}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}\left(u_t^h = \mathrm{pass2}\right) \tag{11}$$

$$\mathrm{RecFrac}^h{}_{\mathrm{tracked}} = \left(\frac{1}{T'}\right) \sum_{t=1}^{T'} \mathrm{I}\left(u_t^h = \mathrm{tracked}\right) \tag{12}$$

These fractions distinguish videos dominated by observed frames from videos where coverage is achieved primarily through temporal inference or short-gap propagation.

Failure mode rates

To characterize dominant degradations, we report per-video failure rates computed from the exported stream. For two-hand failure modes we use a *hand-summed* rate (range $[0, 2]$), so the value can exceed 1 when both hands are affected in the same frame.

Missing rate**.** The fraction of hand-frames where landmarks are missing:

$$\text{MissingRate} = \left(\frac{1}{T'}\right)\sum_{t=1}^{T'}[\mathrm{I}(H_t^L = \emptyset) + \mathrm{I}(H_t^R = \emptyset)] \tag{14}$$

Occlusion rate. The fraction of hand-frames flagged as occluded by the pipeline:

$$\text{OcclusionRate} = \left(\frac{1}{T'}\right)\sum_{t=1}^{T'}[\mathrm{I}(u_t^L = occluded) + \mathrm{I}(u_t^R = occluded)] \tag{15}$$

Swap rate. When left–right identity is ambiguous, the pipeline may apply corrective operations such as swapping left/right assignments. Let S_t be binary indicator of whether a swap is applied at frame *t*. We report the corresponding intervention rates:

$$\text{SwapRate} = \left(\frac{1}{T'}\right)\sum_{t=1}^{T'} S_t \tag{16}$$

Geometric outlier rate. To summarize temporal instability, we flag frames where an implementation-aligned criterion detects abnormally large normalized inter-frame changes in hand configuration and/or apparent scale. Let O_t^L and O_t^R be binary indicators of whether an outlier is flagged for the left and right hand at frame t. For failure-mode profiling, we report the hand-summed per-video outlier rate:

$$\text{OutlierRate} = \left(\frac{1}{T'}\right)\sum_{t=1}^{T'}[O_t^L + O_t^R] \tag{17}$$

Together, these diagnostics complement the composite score by capturing how often the pipeline must infer hands beyond primary detection, how frequently identity corrections are triggered, and whether the resulting trajectories are temporally stable properties that directly affect the usefulness of skeleton streams for SLR training.

Results. This section reports dataset-scale reliability of the proposed fault-tolerant keypoint extraction pipeline on Slovo (*n* = 20,400 clips). The analysis targets stream robustness (availability and stability of exported pose–hand keypoints) rather than downstream recognition accuracy. Using the metrics defined in Section 5, we summarize: **(i)** hand-keypoint availability over time via any-hand coverage Cov_{any} (Eq. 6), per-hand coverages Cov_L, Cov_R(Eqs. 4–5), and both-hands coverage Cov_{LR} (Eq. 7); **(ii)** the distribution of the composite quality score Q used for ranking and filtering (Eq. 10); **(iii)** failure-mode rates - $MissingRate$ (Eq. 14), $OcclusionRate$ (Eq. 15), $SwapRate$ (Eq. 16), and $OutlierRate$ (Eq. 17) - reported with tail-sensitive statistics; and **(iv)** the contribution of recovery mechanisms via per-hand recovery fractions (Eqs. 11–13).

Overall, the goal is to provide a compact, reproducible quality report for monitoring, dataset filtering, and prioritizing engineering improvements, supported by an ablation comparison against a single-pass baseline.

Dataset coverage

We first summarize how consistently the pipeline exports hand keypoints over time. We report any-hand coverage Cov_{any} (Eq. 6), defined as the fraction of frames in which at least one hand is present, alongside per-hand coverages Cov_L, Cov_R (Eq. 4–5) and both-hands coverage Cov_{LR} (Eq. 7).

At the dataset scale, Cov_{any} is high for most clips: the 10th percentile of Cov_{any} is 0.923, and 70.6% of videos achieve perfect any-hand coverage ($Cov_{any} = 1.0$). Low-coverage cases form a relatively small tail: 7.5% of clips have $Cov_{any} < 0.90$, and severe failures are rare (1.2% with $Cov_{any} < 0.50$). Overall, this indicates that the extraction pipeline is stable on the majority of samples, while a small subset of clips remains challenging (e.g., hands leaving the frame, motion blur, or strong occlusions).

Next, we assess suitability for two-handed modeling using both-hands coverage Cov_{LR} (Eq. 7), the fraction of frames where both hands are simultaneously available. The distribution is strongly bimodal: 20.1% of clips have $Cov_{LR} = 0$ (both hands are never present at the same time), while 20.4% have $Cov_{LR} = 1$ (both hands are always present). For the remaining clips, simultaneous two-hand visibility is generally high (median P_{50}= 0.854, P_{75}= 0.978). However, the existence of a substantial $Cov_{LR} = 0$ group suggests that a non-trivial portion of the dataset is structurally one-handed due to framing or signer behavior. Therefore, downstream experiments that assume two visible hands should either filter by Cov_{LR} or rely on reliability-aware handling of missing parts.

Composite quality score distribution

We next summarize the distribution of the composite quality score Q defined in Eq. (10). Figure 5 reports the empirical CDF of Q over the whole dataset. The distribution is strongly skewed toward high scores: the 25th percentile is 0.598, the median is 0.918, and the 75th percentile is 0.984. This indicates that the majority of clips yield consistently usable pose–hand streams under the proposed extraction configuration, with a large mass concentrated near $Q \approx 1$ (see the right-tail zoom).

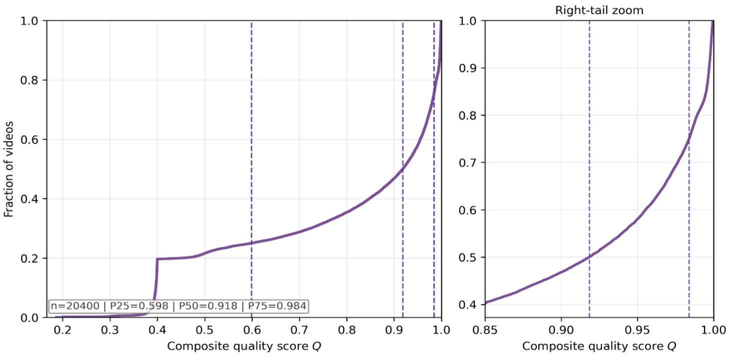

Figure 5. Empirical CDF (ECDF) of the composite quality score Q over Slovo (n = 20,400). Dashed lines indicate P_{25}, P_{50}, and P_{75}; the right panel shows a zoom into the high-score tail.

At the same time, the lower quartile forms a clearly separated low-quality region (below ≈ 0.6), where clips are more likely to suffer from reduced coverage, low confidence, or reliance on recovery mechanisms. Practically, this motivates quality-aware filtering and monitoring: Q can be used (i) as a dataset-level health indicator across runs, and (ii) as a per-clip gating signal to exclude or down-weight low-quality samples in downstream training without requiring ground-truth landmarks.

Failure mode profile

We analyze dominant degradation patterns in the exported pose-hand streams using the per-video failure rates defined in Section 5: MissingRate (Eq. 14), OcclusionRate (Eq. 15), SwapRate (Eq. 16), and OutlierRate (Eq. 17). To emphasize rare but severe failures, Figure 6 reports complementary empirical CDFs (CCDFs), S(x)=P(rate>x)=1−F(x), with a logarithmic y-axis. MissingRate, OcclusionRate, and OutlierRate are computed in hand-summed form (range [0,2]), so values above 1 are possible when both hands are affected in the same frame. SwapRate is bounded to [0,1].

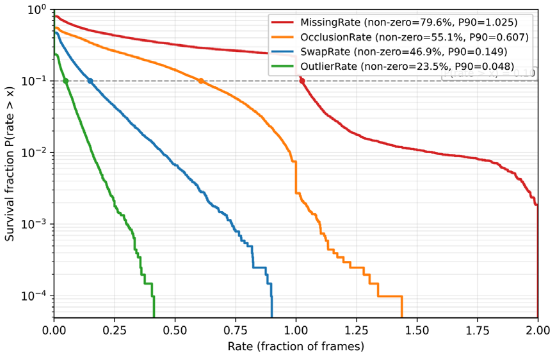

Figure 6. CCDFs (log y-axis) of per-video failure rates on Slovo (n=20,400): MissingRate, OcclusionRate, SwapRate, and OutlierRate. Missing/Occlusion/Outlier are hand-summed ([0,2]); Swap is in [0,1]. Legend shows non-zero clip fraction and P_{90}.

The **results** indicate that missing detections are the primary failure source at dataset scale. MissingRate is non-zero in 79.6% of clips and exhibits the heaviest tail among all modes (P_{90}=1.025), meaning that in the worst decile the stream contains sustained missing landmarks at roughly one affected hand-frame per frame on average (or higher when both hands are missing jointly**).** Occlusions represent the second most impactful failure mode, being non-zero in 55.1% of clips with a substantial tail (P_{90}=0.607), but consistently less severe than missing detections in the high-rate region.

In contrast, identity correction and geometric instability are secondary contributors. SwapRate is non-zero in 46.9% of clips, yet its magnitude remains relatively limited (P_{90}= 0.149), suggesting that identity ambiguities occur but rarely dominate the temporal extent of a clip. OutlierRate is the least prevalent mode (23.5% non-zero) and has the lightest tail (P_{90}= 0.048), consistent with sporadic instability events rather than systematic failure.

Overall, the failure-mode CCDFs provide a clear prioritization for further robustness improvements: the largest dataset-level gains are expected from reducing missing detections and improving occlusion handling, while swap and outlier mitigation primarily target narrower edge cases.

Recovery contribution

We quantify how much of each exported hand stream is produced by recovery mechanisms rather than the primary detector, using the per-clip recovery fractions defined in Eqs. (11), (12). Figure 7 compares the contribution of ROI-based second-pass recovery and tracker-based temporal bridging for left and right hands.

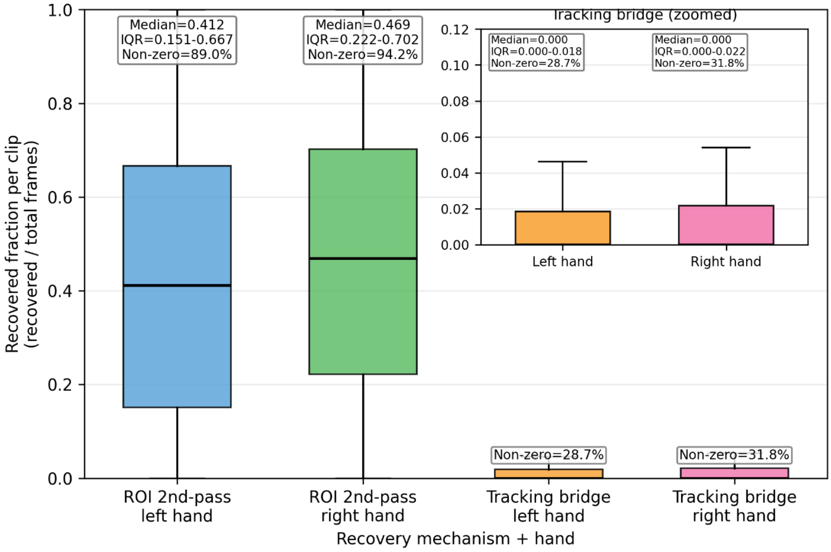

Figure 7. Per-clip recovered fraction (recovered / total clip frames) for ROI second-pass recovery and tracking bridge, separately for left and right hand, on Slovo (n = 20,400). Boxes show median and interquartile range (IQR); inset zooms the tracking regime.

Second-pass recovery is the dominant source at dataset scale: the median recovered fraction is 0.412 for the left hand (IQR 0.151–0.667; non-zero in 89.0% of clips) and 0.469 for the right hand (IQR 0.222–0.702; non-zero in 94.2% of clips). In contrast, tracking contributes little in most clips: the median is 0.000 for both hands, with small upper tails (IQR 0.000–0.018 for left; 0.000–0.022 for right) and non-zero rates of 28.7% and 31.8%, respectively. Overall, robustness is primarily driven by ROI-based second-pass recovery, while tracking acts as a sparse fallback mechanism for a minority of clips.

Ablation study

To quantify the impact of the fault-tolerant components, we compare a single-pass baseline (pass1-only) against the full pipeline on Slovo (n = 20,400 clips). We report a compact set of dataset-level reliability proxies defined in Section 5: the composite quality score Q (Eq. 10), any-hand coverage Cov_{any} (Eq. 6), both-hands coverage Cov_{LR} (Eq. 7), and the missing rate diagnostic MissingRate (Eq. 14).

Table 2 — Ablation results. pass-1 baseline vs. full fault-tolerant pipeline.

Metric	**Pass-1**	**Full**	**Δ**
Composite quality score Q	0.684	0.794	+0.110
Cov_{any}	0.889	0.969	+0.080
Cov_{LR}	0.454	0.646	+0.192
MissingRate	0.329	0.193	−0.136

As shown in Table 2, the full pipeline improves reliability across all reported criteria. The composite score increases by $+0.110$, consistent with a systematic shift toward more usable pose–hand streams. Coverage gains are most pronounced for simultaneous two-hand availability: Cov_{LR} rises by $+0.192$, indicating substantially more frames with both hands present - critical for two-handed signing segments. Any-hand coverage also improves ($+0.080$), while the missing-rate decreases by -0.136, confirming fewer missing hand-frames overall. Collectively, these results support using the full fault-tolerant configuration as the default for large-scale keypoint export and downstream quality-aware filtering.

Discussion. This work argues that pose–hand keypoint extraction for sign-language corpora should be treated as an auditable, quality-controlled process rather than a silent preprocessing step. The proposed quality report operationalizes this view by providing dataset-scale reliability indicators that remain computable without ground-truth landmarks. The Results section shows that the exported streams are robust for most clips, while a clearly identifiable tail concentrates the cases that require filtering, remediation, or pipeline improvements.

A first practical observation is that availability is high but not uniform, and two-hand availability is structurally dataset-dependent. Any-hand coverage Cov_{any} is near-perfect for the majority of clips, whereas both-hands coverage Cov_{LR} is strongly bimodal, with a substantial fraction of clips exhibiting $Cov_{LR} = 0$. This is not necessarily an extractor failure; it often reflects framing and signer behavior and implies that downstream pipelines assuming two visible hands should explicitly gate by Cov_{LR} or adopt reliability-aware handling of missing limbs. In this sense, the coverage metrics act as a data characterization tool in addition to being a robustness indicator.

Second, the distribution of the composite score Q is strongly skewed toward high values, but the lower quartile forms a meaningful low-quality region. This separation supports the use of Q as a compact quality-aware control signal: (i) for monitoring extraction health across runs and configurations, and (ii) for dataset filtering or sample reweighting in downstream training. Importantly, Q is designed to be reproducible under fixed weights (Eq. 10) and interpretable through its components (coverage, pose availability, and confidence), which reduces the risk of treating it as an opaque scalar.

Third, the failure-mode CCDFs provide a clear prioritization for further robustness improvements. Missing detections and occlusions dominate both in prevalence and in tail severity, while identity correction activity and geometric outliers are secondary contributors at the dataset scale. This indicates that the expected return on engineering effort is highest for mechanisms that directly reduce missing intervals and improve occlusion handling (e.g., better hand presence modeling, occlusion-aware refinement, or data-driven thresholds), whereas swap/outlier mitigation targets narrower edge cases.

Fourth, the recovery analysis shows that robustness gains are primarily driven by ROI-based second-pass refinement, with tracking acting as a sparse fallback. Together with the ablation results, this supports a concrete design implication: if compute or implementation complexity must be budgeted, prioritizing a strong second-pass refinement yields most of the dataset-scale benefit, while tracking provides smaller incremental gains and primarily addresses a minority of clips.

The reported metrics are proxy measures of reliability rather than geometric correctness, as no ground-truth landmarks are available. Recovery mechanisms can produce plausible trajectories that may still be inaccurate, so high coverage alone should not be interpreted as guaranteed correctness. Additionally, the diagnostic rates depend on implementation-aligned criteria (e.g., occlusion flags, outlier tests, identity interventions), and their absolute values may shift under domain changes such as viewpoint, resolution, cropping, and motion blur. Finally, while fixed weights make Q comparable across runs, the chosen weighting is still a modeling choice; alternative weightings could be better aligned with particular downstream objectives.

A natural next step is to validate how these reliability indicators correlate with downstream SLR performance (e.g., WER/accuracy vs. Cov_{LR}, $\mathrm{MissingRate}$, and Q). Partial annotation of a small subset would enable calibration of proxy metrics against geometric error and help quantify when recovered frames become harmful. Another direction is adaptive configuration across domains (learned thresholds, domain-conditioned weights for Q, or uncertainty-aware scoring) to improve transferability without sacrificing interpretability. Finally, targeted improvements for missing/occlusion handling are likely to deliver the largest dataset-scale gains, as suggested by the tail behavior in the failure-mode CCDFs.

Conclusion. We presented a quality-aware pose–hand keypoint extraction pipeline for skeleton-based SLR that integrates pose-guided left/right assignment, ROI-based second-pass recovery, short-gap temporal bridging (tracking/hold), sanity validation, and anchor-preserving postprocessing. Crucially, the pipeline exports explicit reliability semantics together with a proxy-metric quality report, turning keypoint extraction into an auditable and reproducible stage that supports dataset-level monitoring, filtering, and reliability-aware downstream training.

Dataset-scale evaluation on Slovo (n = 20,400) shows that the exported streams are usable for the majority of clips: any-hand coverage is near-saturated for most videos, and the composite quality score distribution is strongly concentrated toward high values (P_{25}= 0.598, P_{50}= 0.918, P_{75}= 0.984). Failure-mode profiling indicates that missing detections and occlusions dominate both prevalence and tail severity, making them the primary targets for future robustness gains, while identity correction activity and geometric outliers are comparatively secondary. Recovery attribution further demonstrates that robustness is driven mainly by ROI second-pass refinement (median recovered fraction 0.412/0.469 with non-zero rates 89.0–94.2%), whereas tracking contributes marginally (median ≈ 0 with non-zero rates ≈ 30%), acting primarily as a fallback. Finally, ablation confirms that the full fault-tolerant configuration improves dataset-level reliability across coverage and quality indicators relative to a single-pass baseline, supporting its use as the default setting for large-scale keypoint export and subsequent quality-aware filtering.

References

Bazarevsky V., Grishchenko I., Raveendran K., Zhu T., Zhang F., & Grundmann M. (2020) *BlazePose: On-device real-time body pose tracking.* arXiv:2006.10204. doi:10.48550/arXiv.2006.10204 (in Eng.)

Cao Z., Simon T., Wei S.-E., & Sheikh Y. (2017) *Realtime multi-person 2D pose estimation using part affinity fields.* In CVPR. — P. 7291–7299. doi:10.1109/CVPR.2017.143 (in Eng.)

Duarte A., Palaskar S., Ventura L., Ghadiyaram D., DeHaan K., Metze F., Torres J., & Giro-i-Nieto X. (2021) *How2Sign: A large-scale multimodal dataset for continuous American Sign Language.* In CVPR. — P. 2735–2744. doi:10.1109/CVPR46437.2021.00276 (in Eng.)

Grishchenko I., Bazarevsky V., Zanfir A., Bazavan E.G., Zanfir M., Yee R., Raveendran K., Zhdanovich M., Grundmann M., & Sminchisescu C. (2022) *BlazePose GHUM Holistic: Real-time 3D human landmarks and pose estimation.* arXiv:2206.11678. doi:10.48550/arXiv.2206.11678 (in Eng.)

Jiang S., Sun B., Wang L., Bai Y., Li K., & Fu Y. (2021) *Skeleton aware multi-modal sign language recognition.* In CVPR Workshops. — P. 3413–3423. doi:10.1109/CVPRW53098.2021.00380 (in Eng.)

Kapitanov A., Kvanchiani K., Nagaev A., & Petrova E. (2023) *Slovo: Russian Sign Language dataset.* In Computer Vision Systems (LNCS). — P. 63–73. Springer. doi:10.1007/978-3-031-44137-0_6 (in Eng.)

Kindiroglu A.A., Ozdemir O., & Akarun L. (2019) *Temporal accumulative features for sign language recognition.* In ICCV Workshops. — P. 1288–1297. doi:10.1109/ICCVW.2019.00164 (in Eng.)

Koller O., Forster J., & Ney H. (2015) *Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers.* Computer Vision and Image Understanding, 141. — P. 108–125. doi:10.1016/j.cviu.2015.09.013 (in Eng.)

Koller O., Zargaran S., Ney H., & Bowden R. (2018) *Deep Sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs.* International Journal of Computer Vision, 126. — P. 1311–1325. doi:10.1007/s11263-018-1121-3 (in Eng.)

Li D., Rodriguez C., Yu X., & Li H. (2020) *Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison.* In WACV. — P. 1459–1469. doi:10.1109/WACV45572.2020.9093512 (in Eng.)

Liu J., Shahroudy A., Xu D., & Wang G. (2016) *Spatio-temporal LSTM with trust gates for 3D human action recognition.* In B. Leibe et al. (Eds.), ECCV 2016 (LNCS 9907. — P. 816–833. Springer. doi:10.1007/978-3-319-46487-9_50 (in Eng.)

Lugaresi C., Tang J., Nash H., McClanahan C., Uboweja E., Hays M., Zhang F., Chang C.-L., Yong M.G., Lee J., Chang W.-T., Hua W., Georg M., & Grundmann M. (2019) *MediaPipe: A framework for building perception pipelines.* arXiv:1906.08172. doi:10.48550/arXiv.1906.08172 (in Eng.)

Moryossef A., Tsochantaridis I., Aharoni R., Ebling S., & Narayanan S. (2020) *Real-time sign language detection using human pose estimation.* In Computer Vision – ECCV 2020 Workshops. Springer. doi:10.1007/978-3-030-66096-3_17 (in Eng.)

Moryossef A. (2024) *Optimizing hand region detection in MediaPipe Holistic full-body pose estimation to improve accuracy and avoid downstream errors.* arXiv:2405.03545. doi:10.48550/arXiv.2405.03545 (in Eng.)

Paoletti G., Beyan C., & Del Bue A. (2023) *SKELTER: Unsupervised skeleton action denoising and recognition using transformers.* Frontiers in Computer Science, 5, 1203901. doi:10.3389/fcomp.2023.1203901 (in Eng.)

Song Y.-F., Zhang Z., Shan C., & Wang L. (2021) *Richly activated graph convolutional network for robust skeleton-based action recognition.* IEEE TCSVT, 31(5). — P. 1915–1925. doi:10.1109/TCSVT.2020.3015051 (in Eng.)

Tunga A., Nuthalapati S.V., & Wachs J. (2021) *Pose-based sign language recognition using GCN and BERT.* In WACV Workshops. doi:10.1109/WACVW52041.2021.00008 (in Eng.)

Xing Y., & Zhu J. (2021) *Deep learning-based action recognition with 3D skeleton: A survey.* CAAI Transactions on Intelligence Technology, 6(1). — P. 80–92. doi:10.1049/cit2.12014 (in Eng.)

Yan S., Xiong Y., & Lin D. (2018) *Spatial temporal graph convolutional networks for skeleton-based action recognition.* Proceedings of the AAAI Conference on Artificial Intelligence**,** 32(1). — P. 7444–7452. doi:10.1609/aaai.v32i1.12328 (in Eng.)

Zhang F., Bazarevsky V., Vakunov A., Tkachenka A., Sung G., Chang C.-L., & Grundmann M. (2020) *MediaPipe Hands: On-device real-time hand tracking.* arXiv:2006.10214. doi:10.48550/arXiv.2006.10214 (in Eng.)

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*
Editors: *D.S. Alenov, T. Apendiev*
Computer layout: *G.D. Zhadyranova*

Signed for print: June 15, 2026
Format: 70×90 1/16. 26.5 printed sheets. Order No. 2.

«Central Asian Academic Research Center» LLP
Almaty, Shevchenko Street, 28