

Hand Gesture Recognition for Telugu Sign Language using YOLOV11x and Optimization Techniques

B. K. V. P. S. Mahalakshmi¹, S. Anuradha², K Ram Srujan Reddy³, Karumuru Priyanka Reddy⁴, Afra Khan⁵, Tadala Dileep⁶

^{1,3,4,5,6} School of Technology, Dept of CSE, Gitam Deemed To Be University, Hyderabad, Telangana – India

² School of Technology, Dept of CSE, Gitam Deemed To Be University, Visakhapatnam, Andhra Pradesh – India

సంబంధిత రచయిత *: mbolired@gitam.edu

YOLOV11x మరియు ఆప్టిమైజేషన్ టెక్నిక్‌లను ఉపయోగించి తెలుగు సంజ్ఞ భాష కోసం చేతి సంజ్ఞ గుర్తింపు

B. K. V. P. S. Mahalakshmi¹, S. Anuradha², K Ram Srujan Reddy³, Karumuru Priyanka Reddy⁴, Afra Khan⁵, Tadala Dileep⁶

^{1,3,4,5,6} School of Technology, Dept of CSE, Gitam Deemed To Be University, Hyderabad, Telangana – India

² School of Technology, Dept of CSE, Gitam Deemed To Be University, Visakhapatnam, Andhra Pradesh – India

సంబంధిత రచయిత *: mbolired@gitam.edu

వియక్త

సంకేత భాష చాలా అవసరం, ఇది దృశ్య వ్యక్తీకరణ పద్ధతిని అందిస్తుంది. అయితే, ప్రస్తుత వ్యవస్థలు ప్రధానంగా అమెరికన్ మరియు భారతీయ సంకేత భాష వంటి భాషలకు మద్దతు ఇస్తాయి, తెలుగు వంటి ప్రాంతీయ భాషలకు వనరులు లేవు. ఈ అధ్యయనం ఈ అంతరాన్ని తగ్గించడానికి రియల్-టైమ్ హ్యాండ్ సంజ్ఞ గుర్తింపు నమూనాను ప్రతిపాదిస్తుంది, TensorRT, కత్తిరింపు మరియు క్వాంటైజేషన్ వంటి ఆప్టిమైజేషన్ పద్ధతులతో YOLOv11x మోడల్‌ను ఉపయోగించి తెలుగు సంజ్ఞలను టెక్స్ట్ గా మారుస్తుంది. ఈ వ్యవస్థ తెలుగు సంజ్ఞల యొక్క అనుకూల డేటాసెట్‌పై శిక్షణ పొందింది, అధిక ఖచ్చితత్వం మరియు గణన సామర్థ్యాన్ని సాధిస్తుంది, నిజ-సమయ అనువాదాన్ని ప్రారంభిస్తుంది. ఈ

వ్యవస్థ ప్రాప్యత చేయగల కమ్యూనికేషన్ సాధనాలలో, ముఖ్యంగా తెలుగు విస్తృతంగా మాట్లాడే ప్రాంతాలలో సంభాష్య అనువర్తనాలను కలిగి ఉంది.

కీలకపదాలు

తెలుగు సంజ్ఞ భాష (TSL), భారతీయ సంజ్ఞ భాష (ISL), చేతి సంజ్ఞ గుర్తింపు (HGR), YOLOv11x

1. పరిచయం

WHO (ప్రపంచ ఆరోగ్య సంస్థ) ప్రకారం, భారతదేశంలో దాదాపు 63 మిలియన్ల మంది వినికిడి లోపంతో బాధపడుతున్నారు. బధిరుల పాఠశాలల్లో ఎవరూ మాట్లాడలేని వారు లేరని సర్వేలో తెలింది ఎందుకంటే వారు వాతావరణంలో నివసిస్తున్నందున వారు వినలేరు. కాబట్టి, వారు మాట్లాడటానికి ఇష్టపడరు మరియు కమ్యూనికేషన్ కోసం సంజ్ఞా భాషను మాత్రమే ఇష్టపడతారు. ఈ ప్రపంచంలో మనుగడ సాగించాలంటే, వారు విద్యావంతులు కావాలి, ప్రజలతో సంభాషించాలి, సంజ్ఞా భాషను అర్థం చేసుకోవాలి మరియు వారు కూడా ఉపాధి పొందాలి. ఏప్రిల్ 2016 లో, ప్రధానమంత్రి నరేంద్ర మోడీ [1] సాధికారత సాధించడానికి, వికలాంగులను చేర్చడానికి భారతీయ సంజ్ఞా భాషను అభివృద్ధి చేయడానికి ప్రభుత్వం "పూర్తిగా కట్టుబడి ఉంది" అని అన్నారు. వికలాంగులను చేర్చుకుని త్వరలో ఒక కేంద్రం ఏర్పాటు చేయబడుతుంది.

ప్రభావవంతమైన కమ్యూనికేషన్ మాటల ద్వారా మరియు మాటల ద్వారా కాకుండా ఆలోచనలను మార్పిడి చేసుకునే సామర్థ్యంపై ఎక్కువగా ఆధారపడి ఉంటుంది. వినికిడి మరియు ప్రసంగ వైకల్యం ఉన్న వ్యక్తులకు, సంజ్ఞ భాష ప్రాథమిక మాధ్యమంగా పనిచేస్తుంది. అమెరికన్ సంజ్ఞ భాష (ASL) మరియు ఇండియన్ సంజ్ఞ భాష (ISL) వంటి భాషలు సంజ్ఞ గుర్తింపు వ్యవస్థల ద్వారా సాంకేతిక మద్దతును పొందినప్పటికీ, తెలుగు వంటి ప్రాంతీయ భాషలకు మద్దతు ఇచ్చే వ్యవస్థలలో గణనీయమైన అంతరం ఉంది. ఈ అధ్యయనం తెలుగు సంజ్ఞలను టెక్స్ట్ లోకి అనువదించడానికి స్కెలబుల్ హ్యాండ్ సంజ్ఞ గుర్తింపు నమూనాను పరిచయం చేస్తుంది, ఇది తెలుగు మాట్లాడే వ్యక్తులకు స్వతంత్ర కమ్యూనికేషన్ పద్ధతిని అందిస్తుంది. కమ్యూనికేషన్ అనేది మానవ పరస్పర చర్య యొక్క ప్రాథమిక అంశం, మరియు వినికిడి లేదా ప్రసంగ లోపాలు ఉన్న వ్యక్తులకు, సంజ్ఞ భాష వ్యక్తీకరణ యొక్క ప్రాథమిక మార్గంగా పనిచేస్తుంది. తెలుగు మాట్లాడే సమాజం ఉపయోగించే తెలుగు సంజ్ఞ భాష (TSL), చేతి సంజ్ఞలు మరియు శరీర కదలికల ద్వారా కమ్యూనికేట్ చేయడానికి నిర్మాణాత్మక మార్గాన్ని అందిస్తుంది. అయితే, సాధారణ

జనాభాలో TSL గురించి విస్తృతమైన అవగాహన లేకపోవడం చెవిటి మరియు వినికిడి లోపం ఉన్న సమాజానికి ప్రాప్యత మరియు చేరికలో అడ్డంకులను సృష్టిస్తుంది.

పరిశోధన లక్ష్యాలు మరియు పద్ధతి

ఈ ప్రాజెక్ట్ యొక్క ప్రాథమిక లక్ష్యం తెలుగు సంజ్ఞ భాషతో అనుబంధించబడిన సంజ్ఞలను ప్రత్యేకంగా టెక్స్ట్‌లోకి అనువదించే రియల్-టైమ్, స్కేలబుల్ హ్యాండ్ సంజ్ఞ గుర్తింపు వ్యవస్థను అభివృద్ధి చేయడం. లక్ష్యాలు:

- తెలుగు అక్షరమాలకు సంబంధించిన సంజ్ఞలను సమర్థవంతంగా గుర్తించి అనువదించగల నమూనాను నిర్మించడం.
- గణన సామర్థ్యాన్ని పెంచడానికి మరియు వ్యవస్థను రియల్-టైమ్ అనువర్తనాలకు ఆచరణీయంగా మార్చడానికి అధునాతన ఆప్టిమైజేషన్ పద్ధతులను వర్తింపజేయడం.
- ఈ నమూనా అధిక ఖచ్చితత్వం మరియు విశ్వసనీయతతో పనిచేస్తుందని నిర్ధారించుకోవడం, ముఖ్యంగా తెలుగు వర్ణమాల యొక్క ప్రత్యేకతపై దృష్టి సారించడం, ఇది సాధారణంగా అధ్యయనం చేయబడిన భాషలతో పోలిస్తే విభిన్న దృశ్య సవాళ్లను కలిగిస్తుంది.

2. సాహిత్య సర్వే

సంజ్ఞ భాష కోసం చేతి సంజ్ఞ గుర్తింపులో అనేక విధానాలు అన్వేషించబడ్డాయి:

శివశంకర మరియు శ్రీనాథ్[2] అధిక ఖచ్చితత్వంతో స్టాటిక్ సంజ్ఞలపై దృష్టి సారించి, ASL గుర్తింపు కోసం HSV కలర్ డిటెక్షన్ మరియు సెంట్రాయిడ్ ట్రాకింగ్‌ను ఉపయోగించారు. రౌట్, MMD . మరియు ఇతరులు [3] ఓట్సు యొక్క అల్గోరిథం ఉపయోగించి చెవిటి వ్యక్తుల కోసం సంకేతాలను గుర్తించడానికి ఒక పద్ధతిని అభివృద్ధి చేశారు. ప్రతి చేతి సంజ్ఞ చిత్రం వెబ్‌క్యామ్ నుండి సంగ్రహించబడి, దానిని గ్రే స్కేల్ ఇమేజ్‌గా మారుస్తుంది మరియు దాని విభజన ఓట్సు థ్రెషోల్డింగ్ అల్గోరిథం ఉపయోగించి నిర్వహించబడుతుంది మరియు సంబంధిత సంజ్ఞ చిత్రం కోసం అవుట్‌పుట్ టెక్స్ట్‌ను ఉత్పత్తి చేస్తుంది. ఇమేజ్ వర్గీకరణ పనులలో వాటి ప్రభావం కారణంగా కన్వల్యూషనల్ న్యూరల్ నెట్‌వర్క్‌లు (CNN లు) సాధారణంగా సంజ్ఞ గుర్తింపులో వర్తించబడతాయి. అయితే, గణన డిమాండ్ల కారణంగా CNN లు తరచుగా నిజ-సమయ అనువర్తనంలో పరిమితులను ఎదుర్కొంటాయి [4]. ప్రారంభ పద్ధతులు ప్రధానంగా ఫీచర్ ఇంజనీరింగ్ మరియు హ్యాండ్ సంజ్ఞలను వర్గీకరించడానికి హిడెన్ మార్కోవ్ మోడల్స్ (HMM లు) మరియు సపోర్ట్

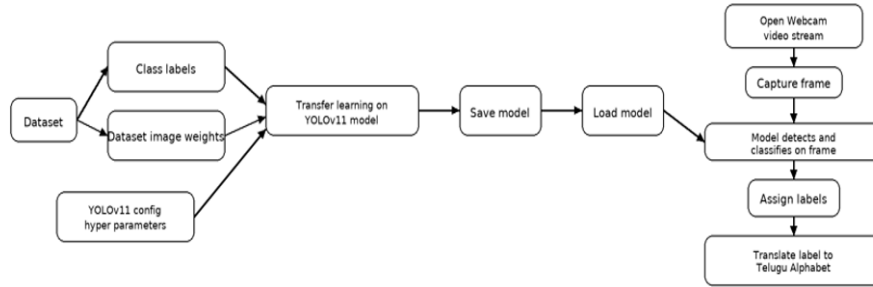
వెక్టర్ మెషిన్స్ (SVM లు) వంటి సాంప్రదాయ వర్గీకరణలపై ఆధారపడి ఉన్నాయి [5], [6]]. అయితే, ఈ పద్ధతులు సంతకం శైలులు, మూసివేతలు మరియు నిజ-సమయ అమలులో వైవిధ్యాలను ఎదుర్కోవడంలో సవాళ్లను ఎదుర్కొన్నాయి.

YOLO మోడల్లు రియల్-టైమ్ ఆబ్జెక్ట్ డిటెక్షన్లో అధిక పనితీరును చూపించాయి. అయితే, YOLO మోడల్లకు గణనీయమైన గణన శక్తి అవసరం, ఇది యాక్సెస్ చేయగల అప్లికేషన్లకు పరిమితులను కలిగిస్తుంది [7].

ఈ అధ్యయనాలు మోడల్ ఖచ్చితత్వం, వేగం మరియు హార్డ్వేర్ అనుకూలతను సమతుల్యం చేయడం యొక్క ప్రాముఖ్యతను నొక్కి చెబుతున్నాయి. తక్కువ-శక్తి పరికరాల్లో నిజ-సమయ పనితీరును సాధించడానికి ఆప్టిమైజేషన్ పద్ధతులతో YOLOv11x[8] ని ఉపయోగించి, తెలుగు వంటి ప్రాంతీయ భాషలకు ఈ విధానాలను స్వీకరించడం ప్రతిపాదిత వ్యవస్థ లక్ష్యం.

3. పద్ధతి

ప్రతిపాదిత తెలుగు సంకేత భాష గుర్తింపు వ్యవస్థ లోతైన అభ్యాస పద్ధతులను రియల్-టైమ్ హ్యాండ్ ల్యాండ్మార్క్స్ తో అనుసంధానిస్తుంది. వెలికితీత కు సాధించు ఖచ్చితమైన మరియు సమర్థవంతమైన వర్గీకరణ. ఈ పద్ధతి నాలుగు ప్రాథమిక దశలను కలిగి ఉంటుంది: డేటా సేకరణ, ప్రీప్రాసెసింగ్, లక్షణం వెలికితీత, మరియు వర్గీకరణ YoloV11X ఉపయోగించి.



చిత్రం: అభివృద్ధి చేయబడిన ప్రతిపాదిత వ్యవస్థ యొక్క నిర్మాణ దృశ్యం

సిస్టమ్ ఆర్కిటెక్చర్, సిస్టమ్కు కనెక్ట్ చేయబడిన కెమెరాను ఉపయోగించి రియల్-టైమ్లో హ్యాండ్ సైగలను క్యాప్చర్ చేయడంతో ప్రారంభమవుతుంది. ప్రతి క్యాప్చర్ చేయబడిన ఫ్రేమ్ YOLOv11x మోడల్ ద్వారా విశ్లేషణ కోసం సిద్ధం చేయడానికి రీసైజింగ్, నార్మలైజేషన్ మరియు ఆగ్మెంటేషన్తో సహా ప్రీప్రాసెసింగ్కు

లోనవుతుంది. ప్రిప్రాసెసింగ్ దశ తర్వాత, ఫ్రేమ్ YOLOv11x మోడల్ ద్వారా పంపబడుతుంది, ఇది రియల్-టైమ్లో సంజ్ఞలను గుర్తించి వర్గీకరిస్తుంది.

గుర్తించిన సంజ్ఞలను తెలుగులోని సంబంధిత వచనానికి మ్యాప్ చేస్తారు, ఇది తెరపై ప్రదర్శించబడే లేదా ఇతర ఇంటర్ఫేస్లకు ప్రసారం చేయగల అనువాదాన్ని అందిస్తుంది. ఈ డేటా ప్రవాహం సంజ్ఞలను గుర్తించడం, ప్రాసెస్ చేయడం మరియు నిరంతర లూప్లో అనువదించడం, నిజ-సమయ అనువాదం యొక్క లక్ష్యాన్ని సాధించడం నిర్ధారిస్తుంది.

4. ప్రయోగాత్మక సెటప్ మరియు అమలు

ఉపయోగించిన సాధనాలు మరియు సాంకేతికతలు

- **ఫ్రేమాన్ :** మోడల్ అభివృద్ధి మరియు ఏకీకరణకు ప్రాథమిక ప్రోగ్రామింగ్ భాష.
- **అట్లాంటిక్స్ YOLO:** YOLOv11x మోడల్ విస్తరణ మరియు శిక్షణ కోసం ఉపయోగించబడుతుంది.
- **ఓపెన్సివి:** రియల్-టైమ్ వీడియో క్యాప్చర్ మరియు ఇమేజ్ ప్రిప్రాసెసింగ్ను సులభతరం చేస్తుంది, మోడల్లోకి డేటాను ఫీడ్ చేయడానికి ఇది అవసరం.
- **టెన్సర్ఫ్లో మరియు పైటోన్:** రెండు ఫ్రేమ్వర్క్లు మోడల్కు శిక్షణ ఇవ్వడానికి మరియు ఫైన-ట్యూన్ చేయడానికి ఉపయోగించబడ్డాయి, ఉత్తమ పనితీరు గల ఫ్రేమ్వర్క్ను ఎంచుకోవడంలో వశ్యతను సమర్థించాయి.
- **పాండాలు మరియు నమ్పీ:** డేటా నిర్వహణ మరియు మానిప్యులేషన్ కోసం నియమించబడ్డారు, ఇది మృదువైన డేటాసెట్ ప్రాసెసింగ్ మరియు ఇంటిగ్రేషన్ను నిర్ధారిస్తుంది.

డేటా సేకరణ: ప్రారంభ డేటా సేకరణలో తెలుగు సంజ్ఞ భాషకు ప్రత్యేకమైన చేతి సంజ్ఞలను సంగ్రహించడం జరిగింది. డేటాసెట్ను విస్తరించడానికి, మోడల్ దృఢత్వాన్ని పెంచడానికి భ్రమణం, స్కేలింగ్ మరియు మొజాయిక్ పరివర్తనాలు వంటి ఆగ్మెంటేషన్ పద్ధతులు వర్తింపజేయబడ్డాయి.

తెలుగు అక్షరాల కోసం అనుకూలీకరించిన డేటాసెట్: సంజ్ఞా భాష సంజ్ఞల వీడియో రికార్డింగ్లను ఉపయోగించి 52 తెలుగు అక్షరాలతో కూడిన డేటాసెట్ను రూపొందించారు. డేటాసెట్ ఉంది క్యూరేటెడ్ కు చేర్చు వైవిధ్యాలు లో లైటింగ్, మోడల్ దృఢత్వాన్ని పెంపొందించడానికి నేపథ్యం మరియు చేతి ధోరణి .

పట్టిక I : అనుకూలీకరించిన తెలుగు డేటాసెట్

అ	ఆ	ఇ	ఈ	ఉ	ఊ	ఋ	ౠ	ఎ	ఏ
ఐ	ఒ	ఓ	ఔ	అం	అః	క	ఖ	గ	ఘ
జ	చ	ఛ	జః	ఝ	ఞ	ట	థ	డ	ఢ
ణ	త	థ	ద	ధ	న	ప	ఫ	బ	భ
మ	య	ర	ల	వ	శ	ష	స	హ	ళ
క్ష	ర								

మోడల్ శిక్షణ మరియు బదిలీ అభ్యాసం : YOLOv11x మోడల్కు సాధారణ డేటాసెట్లపై ముందస్తు శిక్షణ ఇవ్వబడింది, తరువాత తెలుగు సంజ్ఞ డేటాసెట్పై చక్కగా ట్యూన్ చేయబడింది. బ్యాచ్ పరిమాణం మరియు అభ్యాస రేటు వంటి హైపర్పారామీటర్లను గ్రిడ్ శోధన ద్వారా ఆప్టిమైజ్ చేశారు మరియు బదిలీ అభ్యాసం అనుసరణ ప్రక్రియను వేగవంతం చేసింది.

YOLOv11x మోడల్కు శిక్షణ: YOLOv11x మోడల్ డేటాసెట్ నుండి తెలుగు అక్షరమాల సంకేతాలను అధిక ఖచ్చితత్వంతో గుర్తించడం మరియు వర్గీకరించడం నేర్చుకుంటుంది.

YOLOv11x మోడల్లో ట్రాన్స్ఫర్ లెర్నింగ్ : ట్రాన్స్ఫర్ లెర్నింగ్: ఇది ప్రి-ట్రైన్డ్ YOLOv11x మోడల్ను ఉపయోగిస్తుంది. ఇది ప్రి-ట్రైన్డ్ వెయిట్లతో ప్రారంభమవుతుంది మరియు మోడల్ సైన్ ఇమేజ్ డేటాసెట్ లో దాని పారామితులను ఫైన్-ట్యూన్ చేస్తుంది. ఇది శిక్షణ సమయాన్ని గణనీయంగా తగ్గిస్తుంది మరియు అంచు మరియు ఆకార గుర్తింపు వంటి ప్రి-లెర్న్డ్ ఫీచర్లను ఉపయోగించడం ద్వారా ఖచ్చితత్వాన్ని మెరుగుపరుస్తుంది.

మూల్యాంకనం మరియు పరీక్ష : ఖచ్చితత్వాన్ని అంచనా వేయడానికి ప్రెసిషన్, రీకాల్ మరియు సగటు సగటు ప్రెసిషన్ (mAP) వంటి మెట్రిక్‌లను ఉపయోగించి పరీక్ష డేటాపై మోడల్ మూల్యాంకనం చేయబడింది. వివిధ పరికరాల్లో పరీక్షించడం అనుకూలత మరియు పనితీరు స్థిరత్వాన్ని నిర్ధారిస్తుంది.

ఖచ్చితత్వం : అధిక ఖచ్చితత్వం సంజ్ఞల యొక్క ఖచ్చితమైన వర్గీకరణను సూచిస్తుంది, అనువాద లోపాలను తగ్గిస్తుంది.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad 1$$

రీకాల్ : అన్ని సంబంధిత సంజ్ఞలను గుర్తించే మోడల్ సామర్థ్యాన్ని కొలుస్తుంది, పరిపూర్ణతను నిర్ధారిస్తుంది.

$$\text{Recall}(C_i) = \frac{\text{True Positives (TP)}_{C_i}}{\text{True Positives (TP)}_{C_i} + \text{False Negatives (FN)}_{C_i}} \quad 2$$

mAP@50: అబ్జెక్ట్ డిటెక్షన్ లేదా సంజ్ఞ గుర్తింపులో, mAP@50 అనేది IoU థ్రెషోల్డ్ 0.5 వద్ద సగటు సగటు ఖచ్చితత్వంగా నిర్వచించబడింది. ఈ మెట్రిక్ సాధారణంగా మోడల్ పనితీరు యొక్క మూల్యాంకనంలో ఉపయోగించబడుతుంది, ప్రత్యేకించి పని చిత్రాలలో వస్తువులు లేదా సంజ్ఞలను గుర్తించడం అయితే.

$$\text{mAP@50} = \frac{1}{C} \sum_{i=1}^C \text{AP}_i \quad 3$$

సాధారణంగా, AP అనేది ప్రెసిషన్-రీకాల్ వక్రరేఖతో అంచనా వేయబడుతుంది మరియు సాధారణంగా సగటు ప్రెసిషన్ అనేది వక్రరేఖ కింద ఉన్న ప్రాంతం లేదా అనేక స్థాయిల రీకాల్‌లో సగటు ఖచ్చితత్వం.

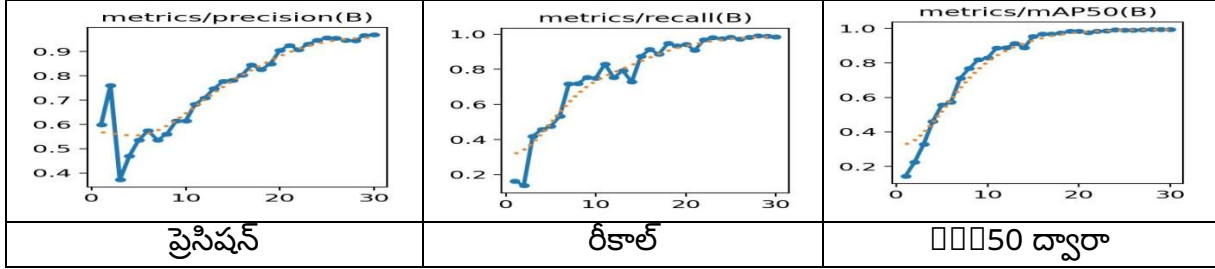
5. ఫలితాల విశ్లేషణ

ఈ ఫలితాలు తెలుగు సంజ్ఞలను ఖచ్చితంగా గుర్తించడంలో, అధిక ఖచ్చితత్వం మరియు జ్ఞాపకశక్తితో మోడల్ యొక్క ప్రభావాన్ని ప్రదర్శిస్తాయి.

పట్టిక II. పనితీరు కొలమానాలు

మోడల్	ప్రెసిషన్	రీకాల్	mAP@50-
యోలోవ్11x	95.97%	94.52%	97.8%

చిత్రం 1. పనితీరు విశ్లేషణ



ఈ ఫలితాలు YOLOv11x తెలుగు సంజ్ఞ భాషలో స్టాటిక్ మరియు డైనమిక్ సంజ్ఞలను గుర్తించడానికి సమర్థవంతమైన మరియు ఖచ్చితమైన పరిష్కారాన్ని అందిస్తుందని నిర్ధారించాయి ఎందుకంటే ఖచ్చితత్వం మరియు రీకాల్ మెట్రిక్స్ స్థిరమైన పెరుగుదల ధోరణిని ప్రదర్శిస్తాయి, ఖచ్చితత్వం 96%కి చేరుకుంటుంది మరియు శిక్షణ ముగిసే సమయానికి రీకాల్ 95%కి చేరుకుంటుంది, ఇది కనీస తప్పుడు పాజిటివ్లు మరియు ప్రతికూలతలతో సంజ్ఞలను ఖచ్చితంగా గుర్తించి వర్గీకరించే మోడల్ సామర్థ్యాన్ని సూచిస్తుంది మరియు mAP@50 మెట్రిక్ దాదాపు 97%కి చేరుకుంటుంది, ఇది వివిధ పరిస్థితులు మరియు సంక్లిష్టతలలో సంజ్ఞలను గుర్తించడంలో మోడల్ యొక్క దృఢత్వాన్ని ప్రతిబింబిస్తుంది. విభిన్న లైటింగ్, నేపథ్యాలు మరియు వినియోగదారు చేతి స్థానాలతో సహా విభిన్న దృశ్యాలను సాధారణీకరించే వ్యవస్థ సామర్థ్యం, వాస్తవ-ప్రపంచ అనువర్తనాలకు దాని అనుకూలతను నొక్కి చెబుతుంది. సారాంశంలో, YOLOv11x మోడల్ చేతి సంజ్ఞ గుర్తింపు కోసం స్కేలబుల్, అధిక-పనితీరు గల ఫ్రేమ్వర్క్ను అందిస్తుంది, తెలుగు మాట్లాడే వినికిడి లోపం ఉన్న సమాజం కోసం నిజ-సమయ అనువాద సాధనాలు మరియు సహాయక సాంకేతికతలలో సంభావ్య అనువర్తనాలతో.

ముగింపు

ఈ పరిశోధన అధ్యయనం తెలుగు మాట్లాడే ప్రాంతాలలోని బధిరులు మరియు మూగ వర్గాలకు రియల్-టైమ్ హ్యాండ్ సంజ్ఞ గుర్తింపు ద్వారా కమ్యూనికేషన్ సులభతరం చేయడంలో గణనీయమైన పురోగతిని అందిస్తుంది. YOLOv11x తో ఆప్టిమైజ్ చేయబడిన CNN ని ఉపయోగించడం ద్వారా, ఈ వ్యవస్థ తెలుగు సంకేత భాషను టెక్స్ట్ లోకి

సమర్థవంతంగా అనువదిస్తుంది. ఈ ఆవిష్కరణ సాంప్రదాయ వివరణ పరిమితులను పరిష్కరిస్తుంది, ఎక్కువ చేరిక మరియు ప్రాప్యతను ప్రోత్సహిస్తుంది. భవిష్యత్

మెరుగుదలలు డేటాసెట్‌ను విస్తరించడం మరియు అదనపు ప్రాంతీయ సంకేత భాషలను చేర్చడం లక్ష్యంగా పెట్టుకుంటాయి, ఇది కలుపుకొని కమ్యూనికేషన్‌కు మరింత దోహదపడుతుంది.

సూచనలు

- [1] భారతీయ సంజ్ఞా భాషను అభివృద్ధి చేయడానికి ఒక కేంద్రం ఏర్పాటు చేయబడుతుంది: PM మోడీ, ఇండియా టుడే, ఏప్రిల్ 11, 2016. [ఆన్‌లైన్]. అందుబాటులో ఉంది: <https://www.indiatoday.in/education-today/news/story/modi-317428-2016-04-11>
- [2] శివశంకర, ఎస్., & శ్రీనాథ్, ఎస్. (2018). అమెరికన్ సంకేత భాష గుర్తింపు వ్యవస్థ: ఒక సరైన విధానం. ఇంటర్నేషనల్ జర్నల్ ఆఫ్ ఇమేజ్, గ్రాఫిక్స్ మరియు సిగ్నల్ ప్రాసెసింగ్, 10(8), 18.
- [3] రౌత్, MMD, థోక్, MP, మచ్చలే, MK మరియు హోరా, MJM, 2015. ఓట్సు అల్గోరిథం ఉపయోగించి బధిరుల కోసం భారతీయ సంకేత భాషను గుర్తించే వ్యవస్థ. ఇంటర్నేషనల్ రీసెర్చ్ జర్నల్ ఆఫ్ ఇంజనీరింగ్ అండ్ టెక్నాలజీ (IRJET), 2(01).
- [4] చుంగ్, HY, చుంగ్, YL, & త్సాయ్, WF (2019, ఫిబ్రవరి). లోతైన CNN ఆధారంగా సమర్థవంతమైన చేతి సంజ్ఞ గుర్తింపు వ్యవస్థ. 2019లో IEEE ఇంటర్నేషనల్ కాన్ఫరెన్స్ ఆన్ ఇండస్ట్రియల్ టెక్నాలజీ (ICIT) (పేజీలు 853-858). IEEE.
- [5] రహేజా, జెఎల్, మిశ్రా, ఎ., & చౌదరి, ఎ. (2016). SVM ఉపయోగించి భారతీయ సంకేత భాష గుర్తింపు. నమూనా గుర్తింపు మరియు చిత్ర విశ్లేషణ, 26(2), 434-441.
- [5] కొల్లర్, ఓ., జర్గర్న్, ఓ., నే, హెచ్., & బౌడెన్, ఆర్. (2016). లోతైన సంకేతం: నిరంతర సంకేత భాష గుర్తింపు కోసం హైబ్రిడ్ CNN-HMM. బ్రిటిష్ మెషిన్ విజన్ కాన్ఫరెన్స్ 2016 యొక్క ప్రొసీడింగ్స్‌లో.
- [7] దివాన్, టి., అనిరుద్, జి., & టెంబుర్నే, జెవి (2023). YOLO ఉపయోగించి వస్తువు గుర్తింపు: సవాళ్లు, నిర్మాణ వారసులు, డేటాసెట్‌లు మరియు అప్లికేషన్లు. మల్టీమీడియా టూల్స్ అండ్ అప్లికేషన్స్, 82(6), 9243-9275.
- [8] ఖానమ్, ఆర్., & హుస్సేన్, ఎం. (2024). యోలోవ్11: కీలకమైన నిర్మాణ మెరుగుదలల యొక్క అవలోకనం. arXiv ప్రీప్రింట్ arXiv:2410.17725.

Hand Gesture Recognition for Telugu Sign Language using YOLOV11x and Optimization Techniques

B. K. V. P. S. Mahalakshmi¹, S. Anuradha², K Ram Srujan Reddy³, Karumuru Priyanka Reddy⁴, Afra Khan⁵, Tadala Dileep⁶

^{1,3,4,5,6} School of Technology, Dept of CSE, Gitam Deemed To Be University, Hyderabad, Telangana – India

² School of Technology, Dept of CSE, Gitam Deemed To Be University, Visakhapatnam, Andhra Pradesh - India

Corresponding Author *: mbolired@gitam.edu

Abstract

Sign language is essential for individuals with hearing and speech disabilities, providing a visual method of expression. However, existing systems primarily support languages like American and Indian Sign Language, lacking resources for regional languages such as Telugu. This study proposes a real-time hand gesture recognition model to bridge this gap, converting Telugu gestures into text using the YOLOv11x model with optimization techniques like TensorRT, pruning, and quantization. The system is trained on a custom dataset of Telugu gestures, achieving high accuracy and computational efficiency, enabling real-time translation. The system holds potential applications in accessible communication tools, especially in regions where Telugu is widely spoken.

Keywords:

Telugu Sign Language (TSL), Indian Sign Language (ISL), Hand Gesture Recognition (HGR), YOLOv11x

1. Introduction

As per WHO (World Health Organization), approximately 63 million people are suffering from hearing Impairment in India. In survey, got to know that no one are speechless in deaf schools because as they are living in the environment, they are unable to hear. So, they are not willing to speak and prefer only sign language for communication. To survive in this world, they should be

educated, communicative with people, understand Sign Language, and they must employ as well. In April 2016, Prime Minister Narendra Modi said [1] to achieving empowerment, The government is “fully committed” to develop Indian sign language to include the Disabled Persons. A Centre would soon be established with inclusion of disabled people.

Effective communication relies heavily on the ability to exchange ideas verbally and non-verbally. For individuals with hearing and speech disabilities, sign language serves as a primary medium. While languages like American Sign Language (ASL) and Indian Sign Language (ISL) have gained technological support through gesture recognition systems, there is a significant gap in systems that support regional languages like Telugu. This study introduces a scalable hand gesture recognition model to translate Telugu gestures into text, providing an independent communication method for Telugu-speaking individuals. Communication is a fundamental aspect of human interaction, and for individuals with hearing or speech impairments, sign language serves as a primary mode of expression. Telugu Sign Language (TSL), used by the Telugu-speaking community, provides a structured way to communicate through hand gestures and body movements. However, the lack of widespread understanding of TSL among the general population creates barriers in accessibility and inclusivity for the deaf and hard-of-hearing community.

Research Objectives and Methodology

The primary goal of this project is to develop a real-time, scalable hand gesture recognition system that specifically translates gestures associated with Telugu sign language into text. The objectives include:

- Building a model that can efficiently recognize and translate gestures corresponding to the Telugu alphabet.
- Applying advanced optimization techniques to enhance computational efficiency and make the system viable for real-time applications.

- Ensuring the model operates with high accuracy and reliability, particularly focusing on the uniqueness of the Telugu alphabet, which poses distinct visual challenges compared to more commonly studied languages.

2. Literature Survey

Numerous approaches have been explored in hand gesture recognition for sign language:

Shivashankara and Srinath[2] utilized HSV colour detection and centroid tracking for ASL recognition, focusing on static gestures with high accuracy. Raut, M.M.D. and et. al [3] developed a method for deaf people using Otsu's Algorithm to identify the signs. Each hand gesture image captured from webcam and convert it into gray scale image, and its segmentation is performed using Otsu thresholding algorithm and produce the output text for corresponding sign image. Convolutional Neural Networks (CNNs) are commonly applied in gesture recognition due to their effectiveness in image classification tasks. However, CNNs often face limitations in real-time application due to computational demands [4]. Early techniques primarily relied on feature engineering and conventional classifiers such as Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) to classify hand gestures [[5], [6]]. However, these methods faced challenges in dealing with variations in signing styles, occlusions, and real-time execution.

YOLO models have shown high performance in real-time object detection. However, YOLO models require significant computational power, which poses limitations for accessible applications [7].

These studies underscore the importance of balancing model accuracy, speed, and hardware compatibility. The proposed system aims to adapt these approaches for regional languages like Telugu, using YOLOv11x[8] with optimization techniques to achieve real-time performance on low-power devices.

3. Methodology

The proposed Telugu Sign Language recognition system integrates deep learning techniques with real-time hand landmark extraction to achieve accurate and efficient classification. The methodology consists of four primary stages: data collection, preprocessing, feature extraction, and classification using YoloV11X.

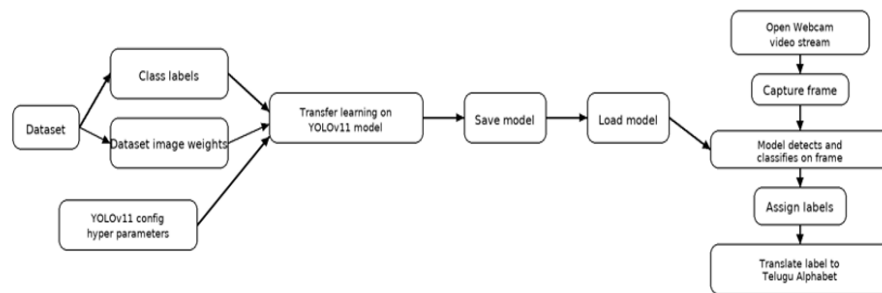


Fig: Architectural view of the Developed Proposed System

The system architecture begins with capturing hand gestures in real-time using a camera connected to the system. Each captured frame undergoes preprocessing, including resizing, normalization, and augmentation, to prepare it for analysis by the YOLOv11x model. After the preprocessing phase, the frame is passed through the YOLOv11x model, which detects and classifies gestures in real-time. The detected gestures are then mapped to the corresponding text in Telugu, providing a translation that can be displayed on-screen or transmitted to other interfaces. This data flow ensures that gestures are detected, processed, and translated in a continuous loop, achieving the objective of real-time translation.

4. Experimental Setup and Implementation

Tools and Technologies Used

- **Python:** The primary programming language for model development and integration.
- **Ultralytics YOLO:** Utilized for YOLOv11x model deployment and training.
- **OpenCV:** Facilitates real-time video capture and image preprocessing, essential for feeding data into the model.

- **TensorFlow and PyTorch:** Both frameworks were used to train and fine-tune the model, supporting flexibility in selecting the best-performing framework.
- **Pandas and NumPy:** Employed for data handling and manipulation, ensuring smooth dataset processing and integration.

Data Collection: Initial data collection involved capturing hand gestures unique to Telugu sign language. Augmentation techniques such as rotation, scaling, and mosaic transformations were applied to expand the dataset, enhancing model robustness.

Customized Dataset for Telugu Alphabets: A dataset comprising 52 Telugu letters was created using video recordings of sign language gestures. The dataset was curated to include variations in lighting, back- ground, and hand orientation to enhance model robustness.

Table I: Customized Telugu Dataset

అ	ఆ	ఇ	ఈ	ఉ	ఊ	ఋ	ౠ	ఎ	ఏ
ఐ	ఒ	ఓ	ఔ	అం	అః	క	ఖ	గ	ఘ
జ	చ	ఛ	జః	రు	రః	ట	ఠ	డ	ఢ
ణ	త	థ	ద	ధ	న	ప	ఫ	బ	భ
మ	య	ర	ల	వ	శ	ష	స	హ	ళ



Model Training and Transfer Learning: The YOLOv11x model was pre-trained on general datasets, then fine-tuned on the Telugu gesture dataset. Hyperparameters, such as batch size and learning rate, were optimized through grid search, and transfer learning accelerated the adaptation process.

Training the YOLOv11x model: The YOLOv11x model learns from the dataset to identify and classify Telugu alphabet signs with high accuracy.

Transfer Learning on YOLOv11x Model: Transfer learning: It uses a pre-trained YOLOv11x model. It begins with pre-trained weights, and the model fine-tunes its parameters on the sign image dataset. This reduces the training time significantly and improves the accuracy by using pre-learned features like edge and shape detection.

Evaluation and Testing: The model was evaluated on test data using metrics such as precision, recall, and mean Average Precision (mAP) to assess accuracy. Testing on different devices ensures compatibility and performance consistency.

Precision: High precision indicates accurate classification of gestures, minimizing translation errors.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad 1$$

Recall: Measures the model's ability to detect all relevant gestures, ensuring completeness.

$$\text{Recall}(C_i) = \frac{\text{True Positives (TP)}_{C_i}}{\text{True Positives (TP)}_{C_i} + \text{False Negatives (FN)}_{C_i}} \quad 2$$

mAP@50: In object detection or gesture recognition, mAP@50 is defined as mean Average Precision at IoU threshold 0.5. This metric is commonly used in the evaluation of a model performance, especially if the task is detecting objects or gestures in images.

$$\text{mAP@50} = \frac{1}{C} \sum_{i=1}^C \text{AP}_i \quad 3$$

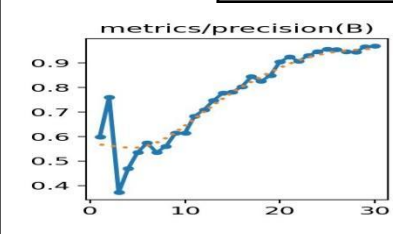
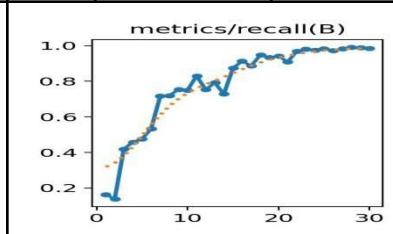
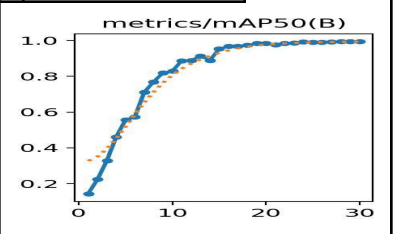
In general, AP is estimated with a precision-recall curve and typically average precision is the area under the curve or an average of precision at several levels of recall.

5. Result Analysis

These results demonstrate the model's effectiveness in accurately recognizing Telugu gestures, with high precision and recall.

Table II. Performance Metrics

Fig1. Analysis	Model	Precision	Recall	mAP@50-Score	Performance
	YOLOV11x	95.97%	94.52%	97.8%	

 metrics/precision(B)	 metrics/recall(B)	 metrics/mAP50(B)
Precision	Recall	mAP50

These results confirm that YOLOv11x provides an efficient and accurate solution for recognizing both static and dynamic gestures in Telugu Sign Language because The precision and recall metrics exhibit a steady upward trend, with precision nearing 96% and recall approaching 95% by the end of training, signifying the model's ability to accurately detect and classify gestures with minimal false positives and negatives and The mAP@50 metric achieves near-perfect performance, reaching approximately 97%, reflecting the model's robustness in detecting gestures across varying conditions and complexities. The system's ability to generalize across diverse scenarios, including different lighting, backgrounds, and user hand positions, underscores its suitability for real-world applications. In summary, the YOLOv11x model offers a scalable, high-performance framework for hand gesture recognition, with potential applications in

real-time translation tools and assistive technologies for the Telugu-speaking hearing-impaired community.

Conclusion

This research study offers a significant advancement in facilitating communication for the deaf and mute communities in Telugu-speaking regions through real-time hand gesture recognition. By employing a CNN optimized with YOLOv11x, the system efficiently translates Telugu sign language into text. This innovation addresses traditional interpretation limitations, promoting greater inclusivity and accessibility. Future enhancements will aim to expand the dataset and incorporate additional regional sign languages, further contributing to inclusive communication.

References

- [1] A centre to be established to develop Indian sign language: PM Modi, India Today, Apr. 11, 2016. [Online]. Available: <https://www.indiatoday.in/education-today/news/story/modi-317428-2016-04-11>
- [2] Shivashankara, S., & Srinath, S. (2018). American sign language recognition system: an optimal approach. *International Journal of Image, Graphics and Signal Processing*, 10(8), 18.
- [3] Raut, M.M.D., Dhok, M.P., Machhale, M.K. and Hora, M.J.M., 2015. A System for Recognition of Indian Sign Language for Deaf People using Otsu's Algorithm. *International Research Journal of Engineering and Technology (IRJET)*, 2(01).
- [4] Chung, H. Y., Chung, Y. L., & Tsai, W. F. (2019, February). An efficient hand gesture recognition system based on deep CNN. In *2019 IEEE International Conference on Industrial Technology (ICIT)* (pp. 853-858). IEEE.
- [5] Raheja, J. L., Mishra, A., & Chaudhary, A. (2016). Indian sign language recognition using SVM. *Pattern Recognition and Image Analysis*, 26(2), 434-441.
- [5] Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2016). Deep sign: Hybrid CNN-HMM for continuous sign language recognition. In *Proceedings of the British Machine Vision Conference 2016*.
- [7] Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6), 9243-9275.
- [8] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.