**A Data Augmentation and Pre-processing Technique for Sign Language Fingerspelling Recognition**

*Fowley, Frank, Ellen Rushe, and Anthony Ventresque. "A Data Augmentation and Pre-processing Technique for Sign Language Fingerspelling Recognition." 24th Irish Machine Vision and Image Processing Conference. 2022.*

**Dataset used:** Synthetic data generated for Irish Sign Language

**Tool** used to generate synthetic data: Blender - greyscale and colour images along with sequences; produces hand joint coordinates, hand bone rotations and palm alignments - along with ground truth labels

The manual elements in ISL include hand shape, orientation, movement and location. - verify if this holds good for ASL as well

**Problems faced** in sign language translation - sparsity of datasets, subject variability, occlusion problems, different lighting and camera perspectives, variations in image resolution and background.

Models where signers are the same for training and testing sets perform significantly better. To address this transfer learning for sign language has been proposed.

Data augmentation is also a potential remedy to the data sparsity challenges as it allows us to create permutations of existing examples to supplement the original dataset.

**Pre-processing techniques used:**

Variations introduced to the dataset - allophonic variations, different camera perspectives, hand-palm size ratio, hand-width to hand-height ratio, minor fluctuations in individual bone rotations, and random rotations of the hand pose about the wrist. - introduced to datasets to improve the accuracy and generalization of models.

Can control signing speed, frames-per-second of videos and image dimensions

The range of variations is constrained to be within anatomical statistical limits reported by Park and Bae, 2020.

Wireframe graphical elements were manipulated to discriminate between similar signs, such as F and G.

*These deliberate geometric and graphical modifications represent a “feature injection,” which aims to add potentially discriminable artificial features into the training dataset samples.*

**Training and testing pipelines:**

Mediapipe was used to extract hand pose landmarks, which were then converted into skeletal wireframe images.

Conversion from RGB pixel images to wireframes alleviates domain adaptation issues such as image background, skin texture and colour.

**Models:**

Learning rate: 0.001

Optimiser: Adam

Batch size: 32 optimal among 16,32, and 64

Model solely trained on the synthetic training dataset

**Results**:

Feature injection - reduction in the image dimensions produced a significant step increase in performance when predicting against the test dataset.

Removal of the palm bones and introduction of depth-coded bone colours provided a large increase in the discrimination between letters.

Model performance is dependent on the accuracy of MediaPipe. MediaPipe had difficulty recognising samples with severely occluded key points.