Supplementary Material for Large Sign Language Models: Toward 3D American Sign Language Translation

A. Implementation details

A.1. Training Hyperparameters

Key hyperparameters for our model training are as follows:

Table 9. Training Hyperparameters

Component	Hyperparameter	Value/Type
Gesture Tokenizer (VQ-VAE)	Codebook Size	1024
	Optimizer	AdamW
	Learning Rate	2×10^{-4}
	Batch Size	256
	Number of Joints	52
LLM + Alignment	Max Token Length	250
	Batch Size per GPU	16
	Alignment MLP LR	1×10^{-6}
	LLM Fine-tune LR	5×10^{-7}
Motion Token IDs	LLaMA	128259
	Qwen	151668

A.2. Dataset Detail

The details of the SignAvatar dataset are listed below:

Table 10. Number of samples in SignAvatars dataset.

Dataset samples	Dev	Test	Val	Language
How2Sign [14]	24476	3060	3059	American Sign Language (ASL)
HamNoSys [26]	4588	574	573	Sign Transcriptive System
RWTH-PHOENIX-Weather [4]	2667	334	333	Germany Sign Language (DGS)

A.3. Dataset Detail

The details of the SignAvatar dataset are listed in List 1.

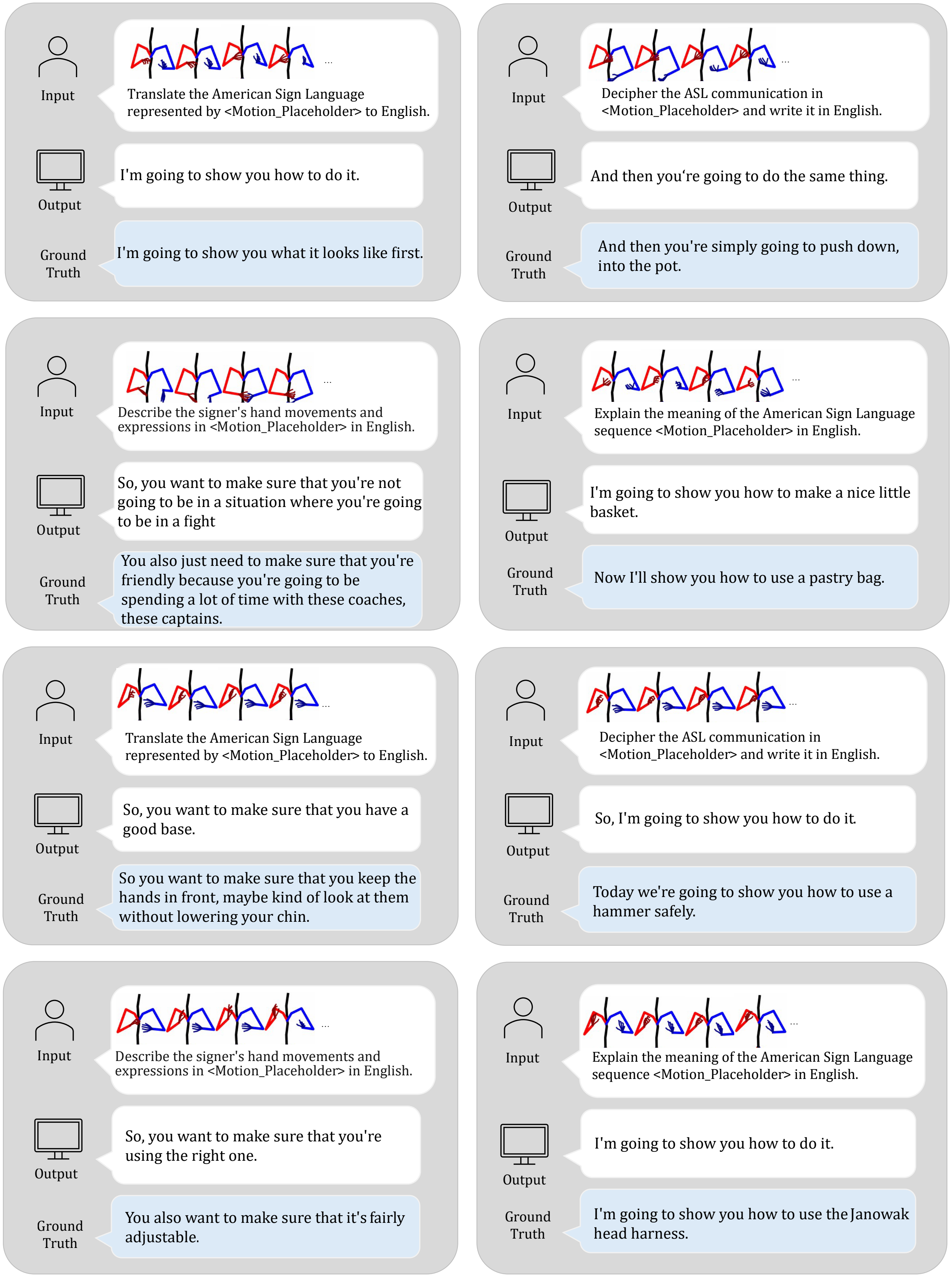
B. More Qualitative Results

We show more results from our LSLM with Qwen backbone in Fig 3.

Listing 1. Templates used for pretraining and instruction-tuning.

```
{
  // Pretraining Template
  "Motion-to-Text": {
    "m2t": {
      "class": "m2t",
      "input": [
        "<Motion_Placeholder>"
      ],
      "output": [
        "<Caption_Placeholder>"
      ]
    }
  }
}

{
  // Instruction-Tuning Template
  "Motion-to-Text": {
    "caption": {
      "class": "m2t",
      "input": [
"Translate the American Sign Language represented by <Motion_Placeholder> to English.",
"Decipher the ASL communication in <Motion_Placeholder> and write it in English.",
"Rephrase the American Sign Language in <Motion_Placeholder> as spoken English.",
"Explain the meaning of the American Sign Language sequence <Motion_Placeholder> in English.",
...
      ],
      "output": [
        "<Caption_Placeholder>"
      ]
    }
  }
}
```



Input

Translate the American Sign Language represented by <Motion_Placeholder> to English.

Output

I'm going to show you how to do it.

Ground Truth

I'm going to show you what it looks like first.

Input

Decipher the ASL communication in <Motion_Placeholder> and write it in English.

Output

And then you‘re going to do the same thing.

Ground Truth

And then you're simply going to push down, into the pot.

Input

Describe the signer's hand movements and expressions in <Motion_Placeholder> in English.

Output

So, you want to make sure that you're not going to be in a situation where you're going to be in a fight

Ground Truth

You also just need to make sure that you're friendly because you're going to be spending a lot of time with these coaches, these captains.

Input

Explain the meaning of the American Sign Language sequence <Motion_Placeholder> in English.

Output

I'm going to show you how to make a nice little basket.

Ground Truth

Now I'll show you how to use a pastry bag.

Input

Translate the American Sign Language represented by <Motion_Placeholder> to English.

Output

So, you want to make sure that you have a good base.

Ground Truth

So you want to make sure that you keep the hands in front, maybe kind of look at them without lowering your chin.

Input

Decipher the ASL communication in <Motion_Placeholder> and write it in English.

Output

So, I'm going to show you how to do it.

Ground Truth

Today we're going to show you how to use a hammer safely.

Input

Describe the signer's hand movements and expressions in <Motion_Placeholder> in English.

Output

So, you want to make sure that you're using the right one.

Ground Truth

You also want to make sure that it's fairly adjustable.

Input

Explain the meaning of the American Sign Language sequence <Motion_Placeholder> in English.

Output

I'm going to show you how to do it.

Ground Truth

I'm going to show you how to use the Janowak head harness.

Figure 4. Instruction-guided Sign Language Recognition (SLR) examples from our LSLM framework with Qwen backbone. Each case shows the model’s translation given a gesture sequence and prompt, alongside the ground truth. Outputs may slightly differ in wording but preserve the core meaning.