Pose-Based Sign Language Appearance Transfer

Amit Moryossef[1, 2*], **Gerard Sant**[1*], **Zifan Jiang**[1]
[1]University of Zurich, [2]sign.mt
amit@sign.mt

Abstract

We introduce a method for transferring the signer's appearance in sign language skeletal poses while preserving the sign content. Using estimated poses, we transfer the appearance of one signer to another, maintaining natural movements and transitions. This approach improves pose-based rendering and sign stitching while obfuscating identity. Our experiments show that while the method reduces signer identification accuracy, it slightly harms sign recognition performance, highlighting a tradeoff between privacy and utility. Our code is available at https://github.com/sign-language-processing/pose-anonymization.

1 Introduction

Personal data, particularly person-identifying information, is central to data protection laws in many countries, including the EU General Data Protection Regulation (GDPR; European Parliament and Council of the European Union (2016)). In signed languages, identifying information is embedded in every utterance through appearance, prosody, movement patterns, and sign choices (Bragg et al., 2020; Battisti et al., 2024). Therefore, from an information-theoretic perspective, removing all identifying information necessitates removing all information. However, a tradeoff between privacy and utility can be achieved by selectively removing some information.

We propose a straightforward yet effective method for altering the appearance of a signer in a sign language pose (Figure 1) while preserving the underlying sign content (§3). Specifically, given a sign language video by signer α and an image of person β, our method generates the appearance of person β performing the same signs as signer α.

Qualitatively, this method effectively smooths skeletal pose stitching (Moryossef et al., 2023b), and improves pose-based video rendering (Saunders et al., 2021). However, quantitative evaluation of our method as data augmentation reveals that while it can help confuse signer identification models, it hurts sign language recognition (§5).

2 Related Work

Research on sign language poses appearance varies in purpose. As Isard (2020) highlights, video anonymization falls into two main categories: concealing parts of the video (Hanke et al., 2020; Rust et al., 2024) or reproducing the video without certain information. This work focuses on the latter.

For instance, Saunders et al. (2021) replace the signer's visual appearance, targeting human consumption. They estimate poses from the original

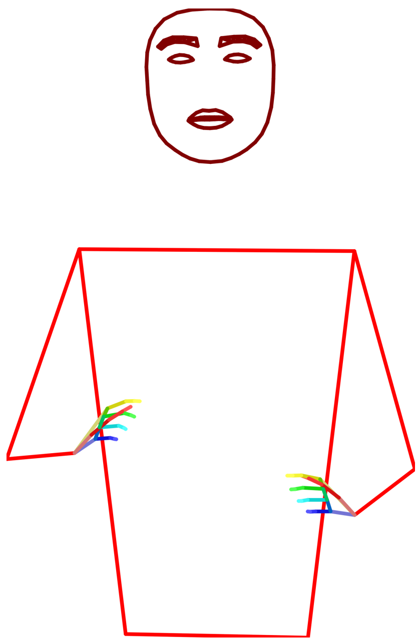


Figure 1: The average MediaPipe Holistic frame (landmarks reduced for visual clarity) extracted from a large sign language dataset (≈ 50 million frames).

Proceedings of the Third International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pages 1–6, June 24, 2025

video and use a Generative Adversarial Network (GAN; Goodfellow et al. (2014)) to generate a different-looking human. This process, working correctly, anonymizes the signing video as effectively as pose estimation alone, since all of the information from the original pose is captured and reproduced. Similarly, cartoon-based anonymization methods replicate signing with animated avatars but often miss key details like facial expressions and hand configurations (Tze et al., 2022).

Battisti et al. (2024) found that pose estimation alone does not conceal signer identity. They noted signers could still be recognized from pose data, highlighting the need for advanced anonymization techniques to better protect privacy. Our work addresses this gap by proposing an appearance transfer to help obfuscate sign language poses.

3 Method

Our appearance transfer approach focuses on altering the appearance of the signer in a pose sequence while preserving the underlying sign information. The method assumes that the video starts from a relaxed posture, not mid-signing.

Given a pose sequence by signer α (P_α), and a single pose frame by signer β (P_β), both poses are normalized to a common scale based on shoulder width, using the `pose-format` (Moryossef et al., 2021a) library. The appearance of both signers is assumed to exist in the first frame of each pose.

Ignoring the hands, to transfer the appearance of signer β to the video by signer α, we modify the pose sequence by removing the appearance of α and adding the appearance of β (Equation 1).

$$\hat{P}_\alpha = P_\alpha - P^0_\alpha + P^0_\beta \tag{1}$$

To perform a standardized anonymization, we choose person β as the mean frame in a large sign language dataset (Figure 1). This results in an average proportioned human, which does not specifically look similar to any individual person. We note that from an information-theoretic perspective, this approach does not guarantee anonymity. Usage is depicted in Algorithm 1.

4 Qualitative Evaluation

This simple approach yields outstanding results. To start, we show a few pose frames from different poses, when transferred to the mean appearance (anonymized) and when transferred to the appearance of a different person (Table 1).

Algorithm 1 'Anonymizing' a pose sequence

```
from pose_format import Pose
from pose_anonymization.appearance \
    import remove_appearance

with open("example.pose", "rb") as f:
  pose = Pose.read(f.read())

pose = remove_appearance(pose)
```

We consider a recent paper on sign language stitching and rendering (Moryossef et al., 2023b). This paper translates spoken language text to sign language videos by identifying relevant signs from a lexicon, stitching them together in a smart way (cropping neutral positions and smoothing the transition), and then rendering a video using a rendering model, trained on a single interpreter. We introduce a single intervention—after finding relevant lexicon items, we transfer the appearance of the pose to be the pose of the interpreter the renderer was trained on.

Rendering The rendering model is a Stable Diffusion model (Rombach et al., 2021) fine-tuned using ControlNet (Zhang and Agrawala, 2023) for controllability from poses. Since the model was trained on the appearance of a single person, it is not robust to various appearances as an input. Generally, it is not a great model, and we would like to maximize the results we get from it. Figure 2 demonstrates the rendering of the face of the original vs. the new pose. We can see that when transferring to the appearance of the interpreter the model was trained on, the results are more 'human'.

(a) Without transfer (b) With transfer

Figure 2: Faces from ControlNet Rendering

Sign Stitching Given a uniform appearance, the stitched pose sequence is now more coherent and less jumpy. The size of different body parts does not change during the sentence, and the stitching points look smoother. When tracking optical flow

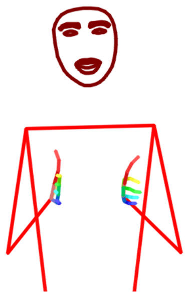
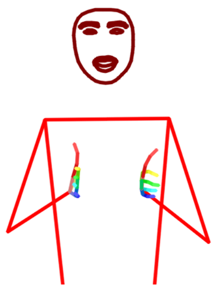
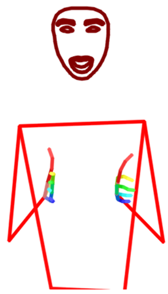
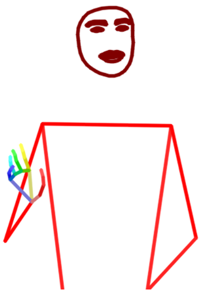
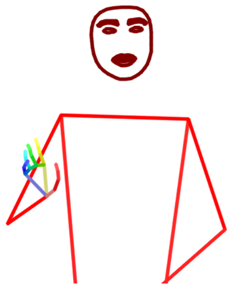
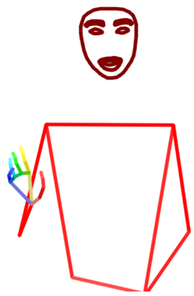
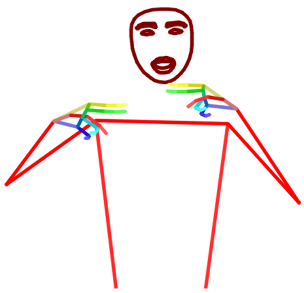
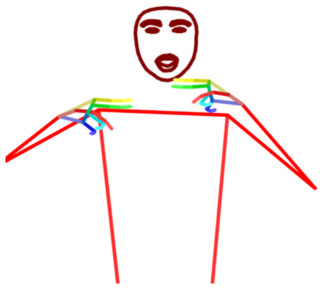
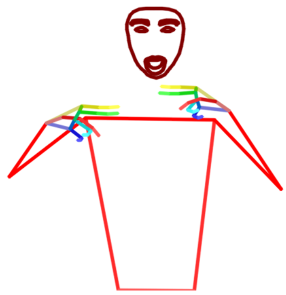
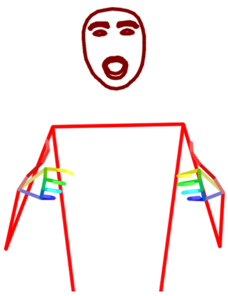
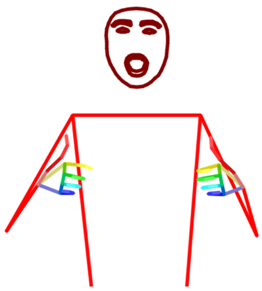
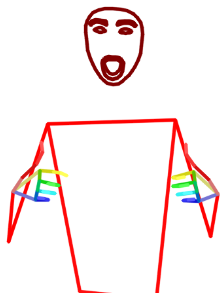
Sign	Original	Anonymized	Transferred
Kleine ('small')			
Kinder ('children')			
essen ('eat')			
Pizza ('pizza')			

Table 1: Example of four signs. On the left, we show the middle frame from the original sign. In the middle, an anonymized version using an average pose from a large sign language dataset. On the right, appearance is transferred to be of a specific interpreter. For a video comparison, check out `https://github.com/sign-language-processing/pose-anonymization`.

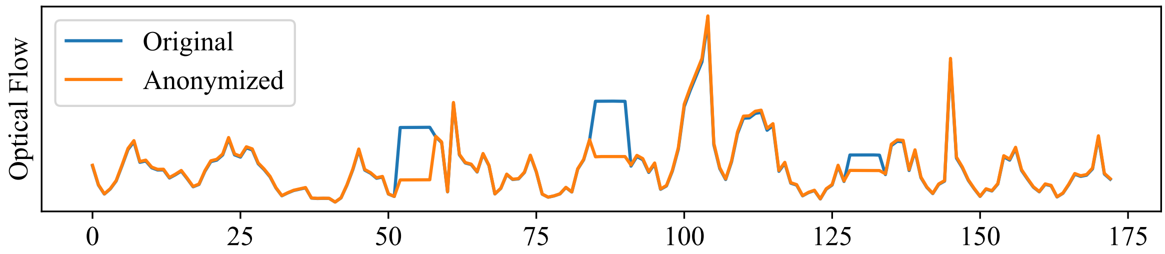


Figure 3: Optical flow (the magnitude of change between two frames) for a stitched video from four original videos and anonymized videos. Higher values represent a larger local change, and a higher area under the curve represents a larger change overall. The flow is exactly the same for all frames except for the stitching zones.

across the pose sequence (Figure 3), sign transitions are smoother and less noticeable, when comparing the use of anonymized and original poses.

5 Experiments and Results

To quantify the effect of our appearance transfer method on sign language recognition, we used the code provided by Moryossef et al. (2021b) for both sign and signer recognition tasks. We hypothesized that transferred poses could serve as an effective data augmentation technique, allowing us to train models to a similar quality while obfuscating signer identities during both training and testing phases.

For our experiments, we used the AUTSL dataset (Sincan and Keles, 2020), which includes 226 distinct lexical sign classes. Importantly, the appearance transfer process did not modify hand pose features, focusing instead on the body and face.

We trained the model under four conditions: (1) using the original pose sequences; (2) applying a single appearance transfer to the average pose shown in Figure 1; (3) transferring multiple appearances for each sample; and (4) combining all these data sources, with 10% original poses, 10% average poses, and 80% transferred appearances. During testing, each model was evaluated on the original pose sequences, transferred to the average pose, and transferred to 10 distinct appearances, with the latter utilizing majority voting, referred to as the *Transferred* method.

As shown in Table 2, no configuration outperformed the model trained and tested with the original pose sequences (top-left). However, training on a combination of original and transferred poses made the model more robust in inference on appearance-augmented data (bottom-right).

To evaluate the extent to which our appearance transfer method obfuscates signer identity, we retrained the model using the original pose sequences but replaced the final sign classification layer with a signer classification layer, freezing the rest of the network as per Sant and Escolano (2023).

Train	**Test**		
	Original	**Anonymized**	**Transferred**
(1) Original Poses	**80.97**%	65.82%	71.46%
(2) Anonymized Poses	63.26%	64.48%	51.50%
(3) Transferred Poses	67.08%	66.54%	57.32%
(4) Combined	79.96%	60.88%	76.78%

Table 2: Sign recognition accuracy on the AUTSL test set. 'Transferred' is an ensemble of predictions from the same 10 different appearances selected randomly.

When trained and tested on the original poses, the model achieved 80.18% accuracy in identifying the signer, demonstrating the existence of identifiable traits. When trained and tested on anonymized poses, accuracy dropped to 65.34%, and with transferred poses, it fell further to 52.20%. These results indicate that while our method significantly reduces identifiable information, it does not eliminate it, as random chance would yield only 3.23% accuracy.

6 Conclusions

We presented a method for appearance transfer in sign language poses, allowing the alteration of a signer's appearance within a pose sequence while preserving essential signing information. By normalizing poses and selectively transferring appearance from another individual—excluding hand geometry to maintain natural movement—we achieved smooth and coherent results in sign rendering and stitching tasks.

Our qualitative evaluation shows that the appearance transfer effectively smooths pose transitions and enhances the visual coherence of stitched sign sequences. However, the quantitative results indicate that while the method helps anonymize signer identity, it can negatively impact sign language recognition performance.

Limitations

We believe that the balance between privacy and utility is to remove all information except for the choice of signs. This is similar to how spoken language text makes speech anonymous to the degree of word choice. Practically, for anonymizing sign language videos, we propose the combination of sign language segmentation (Moryossef et al., 2023a) with phonological sign language transcription. The bottleneck that transcribed sign segments introduce guarantees the removal of identifying information such as appearance, prosodic cues, and movement patterns. Then, a sign language synthesis component should synthesize the transcribed signing sequence back into video.

One major limitation of our study is the lack of human evaluation. While the method aims to preserve essential signing information, it's crucial to assess whether altering the signer's appearance affects the naturalness and comprehensibility of the signs for human viewers, especially in real-world contexts. Evaluating whether the anonymized or transferred appearances still allow viewers to recognize or identify individual signers is key to ensuring the method's success in obfuscating identity. This evaluation will provide insight into how well the technique balances privacy with the utility and intelligibility of the sign content.

Acknowledgements

This work was funded by the SIGMA project (G-95017-01-07) at the Digital Society Initiative (DSI), University of Zurich, and by `sign.mt` ltd.

References

Alessia Battisti, Emma van den Bold, Anne Göhring, Franz Holzknecht, and Sarah Ebling. 2024. Person identification from pose estimates in sign language. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 13–25, Torino, Italia. ELRA and ICCL.

Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring collection of sign language datasets: Privacy, participation, and model performance. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, New York, NY, USA. Association for Computing Machinery.

European Parliament and Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in size and depth. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).

Amy Isard. 2020. Approaches to the anonymisation of sign language corpora. In *SIGNLANG*.

Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, and Yoav Goldberg. 2023a. Linguistically motivated sign language segmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12703–12724, Singapore. Association for Computational Linguistics.

Amit Moryossef, Mathias Müller, and Rebecka Fahrni. 2021a. pose-format: Library for viewing, augmenting, and handling .pose files. `https://github.com/sign-language-processing/pose`.

Amit Moryossef, Mathias Müller, Anne Göhring, Zifan Jiang, Yoav Goldberg, and Sarah Ebling. 2023b. An open-source gloss-based baseline for spoken to signed language translation. In *2nd International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*. Available at: `https://arxiv.org/abs/2305.17714`.

Amit Moryossef, Ioannis Tsochantaridis, Joe Dinn, Necati Cihan Camgöz, Richard Bowden, Tao Jiang, Annette Rios, Mathias Müller, and Sarah Ebling. 2021b. Evaluating the immediate applicability of pose estimation for sign language recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3429–3435.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-resolution image synthesis with latent diffusion models.

Phillip Rust, Bowen Shi, Skyler Wang, Necati Cihan Camgoz, and Jean Maillard. 2024. Towards privacy-aware sign language translation at scale. In *Annual Meeting of the Association for Computational Linguistics*.

Gerard Sant and Carlos Escolano. 2023. Analysis of acoustic information in end-to-end spoken language translation. In *INTERSPEECH 2023*, pages 52–56.

Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. 2021. Anonysign: Novel human appearance synthesis for sign language video anonymisation. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8.

Ozge Mercanoglu Sincan and Hacer Yalim Keles. 2020. Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355.

Christina O. Tze, Panagiotis P. Filntisis, Anastasios Roussos, and Petros Maragos. 2022. Cartoonized anonymization of sign language videos. In *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pages 1–5.

Lvmin Zhang and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models.