

Stable-SCore: A Stable Registration-based Framework for 3D Shape Correspondence

Haolin Liu^{1,2,3*}, Xiaohang Zhan^{2*}, Zizheng Yan^{1,4*}, Zhongjin Luo^{1,4}, Yuxin Wen², Xiaoguang Han^{4,1†}

*equal contribution

†corresponding author

¹FNii, CUHK SZ

²Tencent

³Tencent-Hunyuan3D

⁴SSE, CUHK SZ

haolinliu97.github.io/Stable-Score



Figure 1. We propose Stable-SCore: A Stable Registration-based Framework for 3D Shape Correspondence. Given a source and a target mesh, our approach registers the source mesh to the target and establishes dense correspondence between them, with strong robustness to large variations in mesh topology, shape, and pose, as shown in (a). Our method enables a range of downstream applications, including re-topology, texture transfer, rig transfer, and shape interpolation, as shown in (b).

Abstract

challenging scenarios, and brings possibilities for a wide array of real applications, as demonstrated in our results.

Establishing character shape correspondence is a critical and fundamental task in computer vision and graphics, with diverse applications including re-topology, attribute transfer, and shape interpolation. Current dominant functional map methods, while effective in controlled scenarios, struggle in real situations with more complex challenges such as non-isometric shape discrepancies. In response, we revisit registration-for-correspondence methods and tap their potential for more stable shape correspondence estimation. To overcome their common issues including unstable deformations and the necessity for careful pre-alignment or high-quality initial 3D correspondences, we introduce Stable-SCore: A Stable Registration-based Framework for 3D Shape Correspondence. We first re-purpose a foundation model for 2D character correspondence that ensures reliable and stable 2D mappings. Crucially, we propose a novel Semantic Flow Guided Registration approach that leverages 2D correspondence to guide mesh deformations. Our framework significantly surpasses existing methods in

1. Introduction

Shape correspondence is a fundamental task in computer vision and graphics, where the goal is to establish accurate point-to-point mappings between different shapes, ensuring the preservation of their geometric characteristics. This task involves the identification and alignment of corresponding points, features, or regions across multiple shapes, accommodating variations in pose, scale, or intricate local geometric details. Dense correspondences become especially crucial for 3D characters [11, 12, 33, 41, 68], given its implications for a range of downstream applications such as re-topology, shape interpolation [15, 88], texture transfer [29], rig transfer [58, 71], etc.

There are two primary categories to address shape correspondence: registration-for-correspondence methods [2, 9, 20–22, 40, 70] and functional map methods [7, 11, 16, 17, 27, 42, 54, 57, 65]. In recent years, functional map

methods have dominated this task. These methods transform the challenge of point mapping into one of function mapping, demonstrating leading performance in “controlled” scenarios where discrepancies in shape and mesh topology are small compared to those found in 3D models crafted by artists or generated by AI. Though some of recent work [7, 11, 12] claim applicability to non-isometric settings, their performance falls short when being tested on more challenging non-isometric cases, as shown in Figure 2. This observation makes us rethink the suitability of functional map methods for such tasks. Also highlighted in recent studies [9, 11], functional map methods encounter difficulties with non-isometric correspondence due to their reliance on strictly aligned low-rank basis.

In contrast, the registration-for-correspondence paradigm is essentially more adept at handling non-isometric discrepancies. Given high-quality initial correspondences, these methods [2, 43] iteratively deform the source shape to align it with the target shape, ultimately producing dense correspondences. Despite their straightforward and intuitive approach, registration-for-correspondence methods encounter several challenges. First, existing mesh deformation techniques [2, 20, 21, 31] often result in unstable transformations, suffering from distortion artifacts and frequently struggling to strike a balance between smoothness and deformation accuracy. Secondly, these methods typically depend on careful pre-alignment, or high-quality initial sparse 3D correspondences, which are challenging to secure, particularly when the shapes undergo substantial non-isometric deformations, as depicted in Figure 2. These limitations have hindered the broader adoption and effectiveness of the registration-for-correspondence paradigm.

Embracing registration-for-correspondence as a foundational paradigm, this work reevaluates and addresses its inherent challenges. (1) Initially, we integrate the emerging deformation technique known as Neural Jacobian Fields (NJF) [1] as our deformation model. NJF is recognized for producing stable deformations and is well-suited for applications involving both differentiable rendering and iterative optimization [25, 76, 77]. (2) Given the scarcity of large-scale 3D correspondence datasets, robust initial correspondence estimation seems impractical. However, recent breakthroughs in re-purposing 2D foundational models have demonstrated remarkable capabilities across various 2D tasks such as depth estimation [24, 34, 82] and novel view synthesis [46, 47, 67, 75, 78, 89]. Leveraging this advancement, we train a 2D character correspondence model designed to establish stable 2D correspondences. (3) Crucially, we introduce a novel Semantic Flow Guided Registration framework, equipped with the aforementioned deformation model and the 2D character correspondence model. Within this framework, deformations are

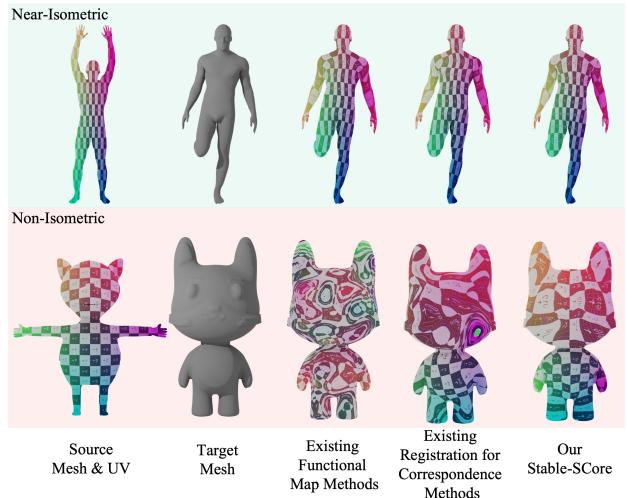


Figure 2. Functional map methods face difficulties with non-isometric correspondences, while existing Registration-for-Correspondence methods break down without reliable initial correspondences. Our Stable-SCore maintains geometric fidelity and offers greater robustness to non-isometric shape variations in the wild.

rendered in a differentiable manner and supervised using 2D correspondences. Via iterative optimization, the source mesh is progressively deformed to align the shape of the target mesh. In this way, this framework produces stable 3D dense correspondence under challenging non-isometric settings.

Furthermore, to enhance robust in-the-wild character correspondence - a practical need in real-world applications - we introduce a challenging new benchmark, *Character in-the-wild (CharW)*. This benchmark consists of meshes from various sources, including professional artists and text-to-3D generation approaches [45, 86]. It features significant geometric and topological variations, capturing a broad spectrum of non-isometric deformations. This diversity ensures that CharW effectively simulates the complex scenarios encountered in practical settings, making it essential for advancing the field of character correspondence.

In summary, our key contributions are as follows:

- We revisit the registration-for-correspondence paradigm, and propose a novel framework, Stable-SCore, to address its key challenges and revitalize its effectiveness.
- We introduce *Character in-the-wild (CharW)*, the first benchmark for character correspondence in-the-wild, featuring shapes with significant geometric and topological variations.
- Through comprehensive experiments, Stable-SCore shows state-of-the-art performance across all non-isometric character correspondence benchmarks, significantly surpassing previous methods.
- We highlight several downstream applications powered

by our approach, such as re-topology, texture transfer, rig transfer, and shape interpolation. These applications showcase the potential of Stable-SCore to open new avenues for a variety of creative and practical uses.

2. Related Work

2.1. 2D Image Correspondence

2D dense correspondence refers to the problem of establishing pixel-wise correspondences between two or more images, based on visual or geometric similarity. Traditional methods often relied on handcrafted feature descriptors such as SIFT [49] or SURF [8]. Recent advancements in deep learning have revolutionized by learning feature representations directly from data [14, 30, 32, 35, 39, 73, 79, 80]. However, due to the limited capacity of the feature extractor and data, their methods struggle with domain gap problem whey applied to in-the-wild images.

Recently, 2D visual foundation models like DINO [56] and Stable Diffusion [63, 64] have demonstrated remarkable generalization capabilities in the image dense correspondence task, outperforming previous 2D correspondence methods [3, 26, 28, 50, 72, 83, 84]. Inspired by [83, 84], we employ a lightweight adapter network that leverages features from Stable Diffusion and DINO to train a 2D character correspondence model.

2.2. 3D Shape Correspondence and Registration

Shape correspondence methods can be broadly categorized into two main approaches: functional map methods and registration methods. The key idea behind functional maps [57] is to express correspondences not as point-to-point matches, but as mappings between functions. FM-Net [16] is the first to combine deep learning and functional map. DiffusionNet [66] proposes an effective spectral feature extractor that pushes the deep functional map methods further. Follow-up works such as [5, 6, 16, 33, 42, 52, 68] extend it and enhance the performance. Some studies [7, 11, 12, 21–23, 27, 33] attempt to address non-isometric correspondence by considering different basis functions or involves extrinsic information. However, they still struggle with the non-isometric setting.

Many axiomatic correspondence techniques rely heavily on registration. Among these, the Non-Rigid ICP [2, 40] are widely used due to their effectiveness in non-rigid registration tasks. [43] propose a point-based hierarchical deformation field for registration. [20] propose a divergence-free deformation field, and alternatively updates correspondence and registration. [9] find correspondences and registers based on control points. [21] proposed a low-rank deformation combining functional map. Recently, some registration-based methods have sought the power of neural networks. [22] uses spectral convolution filters to pro-

cess mesh data, and the output features provide initial correspondence for the registration process as in [21]. [33] use functional map prior for registration through a deformation graph. However, these consistently face some challenges such as unstable deformation processes and require rough alignment or initial correspondence, which limits its application in shape correspondence.

There are some other methods [19, 60–62, 81] who attempt to solve a geometric consistent shape correspondence, in which connection remains after the mapping. [60] propose a 3D-2D mapping method and solve it by finding the shortest path in the product graph. This kind of method shows potential in non-isometric correspondence, however, they are only suitable for low-poly meshes.

2.3. Repurposing 2D Models for 3D Tasks

Recently, 2D vision foundation models have emerged, driven by the availability of billion-scale training datasets. Notable examples include CLIP [59], DINOV2 [56], Stable Diffusion [63], and SAM [36]. In contrast, obtaining datasets of similar scale for building foundation models in 3D vision remains a significant challenge. As a result, many recent works have sought to leverage the powerful generalization capabilities of these 2D foundation models to address 3D tasks, such as depth estimation [24, 34, 82] and novel view synthesis [46, 47, 67, 75, 78, 89]. For instance, Marigold [34] proposes fine-tuning the U-Net of Stable Diffusion for depth estimation by concatenating the input image latent with a noisy latent representation. [69] propose to use Diffmorpher[85] to distill 2D morphing to 3D morphing for shape correspondence. A closely related work, Diff3f [18], back-projects 2D foundational features onto 3D surfaces, and utilize these 3D features for shape correspondences. However, it still requires using the functional map technique to obtain a continuous map. This approach results in a loss of semantic information, as the features are projected onto a low-rank basis, leading to degradation in detail. In contrast, our method re-purposes the foundation model for dense 2D correspondences, with a sophisticated design to guide registration through differentiable rendering, preserving as much detail as possible.

3. Method

3.1. Overview

We propose Stable-SCore: A Stable Registration-based Framework for 3D Shape Correspondence. Given a source mesh \mathcal{M}_{src} and a target mesh \mathcal{M}_{tgt} as inputs, our method first re-purposes 2D foundation models (Sec. 3.2) to extract robust multi-view 2D correspondences. Then, we employ a Semantic Flow-guided Registration method (Sec. 3.3) to guide the registration using the 2D correspondences.

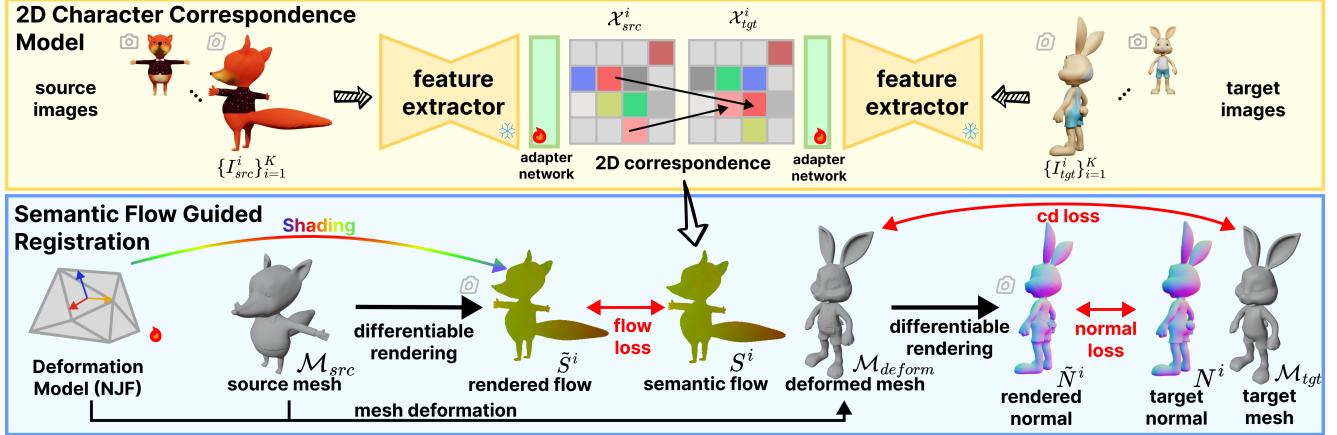


Figure 3. The Stable-SCore pipeline operates as follows: Source and target meshes are inputted and rendered into multi-view RGB or normal images using a fixed set of cameras. These images are processed through the network to extract 2D correspondences as a semantic flow map. Using differentiable rendering, we render forward flow under the same camera views and supervise it with the semantic flow. Chamfer Distance (CD) and normal loss, are also applied. The deformation model is iteratively optimized throughout this process.

3.2. 2D Character Correspondence Model

To achieve accurate and robust 2D correspondences, we adopt the pre-trained foundation models, specifically Stable Diffusion [63] and DINO [56], for correspondence estimation. These models act as feature extractors, followed by an adapter network to extract 2D correspondences, as illustrated in Figure 3.

In detail, for given source and target meshes that are roughly aligned in rotation, we establish a fixed set of camera poses $C_i, i = 1, 2, \dots, K$. Then, we render multi-view images $(I_{src}^i, I_{tgt}^i), i = 1, 2, \dots, K$ for both source and target as input images. The input images can be either rendered as RGB or normal images. We feed these images into Stable Diffusion and DINO to extract feature maps. Specifically, DINO features are collected from the output of DINO, while Stable Diffusion features are extracted from the intermediate layers of the UNet.

Subsequently, a lightweight adapter network is employed to map the source and target features into a common embedding space. Within this space, 2D correspondences are determined through a nearest neighbor search. These correspondences are then encapsulated in a semantic flow map, which illustrates the 2D displacements linking source pixels to their respective target pixels.

In this way, for each camera pose C_i , we find 2D correspondences between I_{src}^i and I_{tgt}^i , producing a semantic flow map S^i .

Training the 2D correspondence model. We train the adapter network while keeping the feature extractors fixed. The network is trainable using both 3D character correspondence and 2D correspondence datasets, including

3DBiCar [51], Surreal [74], and SPair-71K [55]. We believe that with diverse training data, the network is able to estimate robust and accurate 2D correspondences for characters.

During training on 3D character correspondence datasets, we first randomly sample a pair of meshes as the source and the target, and render a pair of images (I_{src}, I_{tgt}) given random camera poses (C_s, C_t) sampled from a prior camera distribution. We force $C_s = C_t$ with a 50% probability. We randomly render normal or RGB images as inputs to train the model. The ground-truth 3D correspondences are projected onto the image plane using (C_s, C_t) to create matched 2D points for supervision. For 2D character correspondence, the pairs of (I_{src}, I_{tgt}) are sampled from the dataset.

(I_{src}, I_{tgt}) are fed into the feature extractor and the adapter network to produce the embeddings $(\mathcal{X}_{src}, \mathcal{X}_{tgt})$. During training, we use a contrastive loss as described in [83]:

$$\mathcal{L}_{con} = CL(\mathcal{X}_{src}(\mathcal{P}^s), \mathcal{X}_{tgt}(\mathcal{P}^t)) \quad (1)$$

where CL is the CLIP-style contrastive loss, \mathcal{P}^s and \mathcal{P}^t are pairs of matched 2D points, $\mathcal{X}_{src}(\cdot)$ and $\mathcal{X}_{tgt}(\cdot)$ denote the processes to sample point features from \mathcal{X}_{src} and \mathcal{X}_{tgt} .

However, relying solely on contrastive loss introduces self-similarity issues, *e.g.*, the left hand being treated similarly to the right hand: different parts may be semantically similar. To address this, we propose a geometry-grounded negative loss. The idea is to reduce self-similarity for a pair of points that are geodesically far away on two meshes respectively. For each pair of points, we are able to compute the pseudo geodesic distance on the common mesh template since we use parametric models such as RaBit [51] and SMPL [48] for training. This loss can be described by

the following equation:

$$\mathcal{L}_{neg} = \sum_{(p,q), \mathbb{G}(p,q) > th} \|\mathcal{X}_{src}(\Pi(p, C_s)) \cdot \mathcal{X}_{tgt}(\Pi(q, C_t))\|_2 \quad (2)$$

where \mathbb{G} is the pseudo geodesic distance matrix pre-computed on the common mesh template, th is the thresholded distance, (p, q) are randomly sampled vertices from the source mesh and target mesh respectively, $\Pi(\cdot)$ denotes the projection operator. The geometry-grounded negative loss is able to alleviate the self-similarity problem and finally achieve more robust shape correspondence. Thus, the total loss to train the 2D character correspondence model is defined as $L_{2D} = L_{con} + \lambda_{neg} L_{neg}$ where λ_{neg} is a weighting factor.

3.3. Semantic Flow Guided Registration

Mesh Deformation Model. Registration-based methods often involve iteratively deforming a source mesh to align it with a target mesh. The mesh deformation model is essential for successful registration, as it aims to produce a smooth deformed result that preserves details while minimizing distortion. We choose Neural Jacobian Fields (NJF) [1] as our deformation model to drive the registration, as it provides powerful and stable deformations, as demonstrated in [25, 76, 77]. Now we introduce some preliminaries of NJF. It first defines per-face Jacobian J_i as follows:

$$J_i \mathcal{B}_i^T [v_k - v_j, v_l - v_j] = [\phi_k - \phi_j, \phi_l - \phi_j] \quad (3)$$

where v_k, v_j, v_l are the original vertices' location of a triangle, ϕ_k, ϕ_j, ϕ_l are the deformed vertices' location, $J_i \in \mathbb{R}^{3 \times 2}$ is the face Jacobian, $\mathcal{B}_i \in \mathbb{R}^{3 \times 2}$ are two-column vectors that form an orthonormal basis for the tangent space of the face. This equation further defines a gradient operator as follows:

$$J_i = \Phi \nabla_i^T, \quad (4)$$

where $\Phi \in \mathbb{R}^{n \times 3}$ are vertices deformed locations. Solving the following Poisson equation yields the deformed vertex locations:

$$\Phi^* = \operatorname{argmin}_{\Phi} \sum_i |t_i| \|\Phi \nabla_i^T - J_i\|_2 \quad (5)$$

Where $|t_i|$ is the face's area. This Poisson equation can be solved in a least-square approach by the following equation:

$$\Phi^* = L^{-1} \mathcal{A} \nabla^T J \quad (6)$$

where L is the mesh's cotangent Laplacian, A is the mesh's mass matrix and J is the stack of estimated Jacobian J_i . During the registration process, we optimize the per-face transformation matrices $\tilde{J}_i \in \mathbb{R}^{3 \times 3}$. These matrices are then projected onto each face's tangent space \mathcal{B}_i to obtain the Jacobian J_i by $J_i = \tilde{J}_i \mathcal{B}_i$. Finally, it applies Equation 6

to solve the final vertices location Φ^* . The intuition behind NJF is to define the deformation in a compact, parameterized tangent space, which simplifies the optimization process compared to performing it in the ambient space. For more details, please refer to [1].

Semantic flow guidance. Given the deformed vertices Φ^* from NJF, we project the vertices and compute the displacement to shade the source mesh \mathcal{M}_{src} and utilize NVDiff [38] to render a differentiable flow map \tilde{S}^i . The shading and rendering procedure can be described as follows:

$$F_i = \Pi(\Phi^*, C_i) - \Pi(\mathcal{V}_{src}, C_i) \quad (7)$$

$$\tilde{S}^i = R(\mathcal{M}_{src}, F_i, C_i) \quad (8)$$

where Φ^* is the deformed vertices location while \mathcal{V}_{src} are source vertices location, C_i is the camera pose same as that used in 2D character correspondence model. We first compute the per-vertex 2D displacement F_i by projecting vertex coordinates into 2D coordinates using the projection operator Π and subtracting them by Equation 7. We further use vertex color shading and assigned the normalized 2D displacement as the vertex color. Finally, a differentiable rendering function $R(\cdot)$ is used to render the flow map \tilde{S}_i under camera pose C_i . The flow loss computed between the multi-view rendered flows and semantic flows S_i yielded from the 2D correspondence model is formulated as follows:

$$\mathcal{L}_{flow} = \sum_{i=1}^K \|\tilde{S}^i - S^i\|_1 \quad (9)$$

Geometry alignment loss. The geometry alignment loss aims at spatially aligning the shape of the deformed source mesh and the target mesh. We apply vertex displacements to obtain the deformed source mesh, \mathcal{M}_{deform} , and compute the Chamfer Distance loss, \mathcal{L}_{cd} , between its vertices and those of the target mesh, \mathcal{M}_{target} . Additionally, we use differentiable rendering to render the normal map of \mathcal{M}_{deform} and compute the normal loss.

Deformation Regularization. To stabilize deformation, we use an identity-preserving term enforced on the per-face transformation matrix as in [25]:

$$\mathcal{L}_{identity} = \sum_i^{|\mathcal{F}|} \|\tilde{J}_i - I_3\|_F \quad (10)$$

However, using only this identity-preserving term is not sufficient. Applying a large weighting factor will impede significant pose deformation while applying a small weight results in unsmooth deformation. To address this problem, we

Table 1. Quantitative comparison between our methods and previous methods. The evaluation metric is mean geodesic error $\times 100$. Ours (Normal) indicates our method with rendered normal images as inputs, aligned with baseline methods that only use geometry, while Ours (RGB) uses rendered RGB images. Ours (Zero-shot) is a variant with adapter network removed and no training is needed. **Bold** indicates the best method while underline indicates the second best.

		Cross Domain				Intra Domain		
		Isometric		Non-isometric		Isometric		Non-isometric
Supervision	Test dataset	FAUST	CharW	DT4D-H std	DT4D-H hard	FAUST	DT4D-H std	DT4D-H hard
Zero-shot	SmoothShell [21]	2.93	11.6	<u>13.6</u>	<u>12.4</u>	2.93	13.6	12.4
	Diff3f [18]	12.0	12.5	24.0	22.7	12.0	24.0	22.7
	Ours (Zero-shot)	5.60	3.48	19.9	14.1	5.60	19.9	14.1
Unsupervised	DeepShell [22]	6.50	37.5	31.0	40.8	1.90	29.1	37.7
	DFR[33]	18.2	<u>6.52</u>	19.8	14.3	9.81	14.9	7.67
	HybridGeoFMap [7]	6.61	32.2	22.1	29.0	2.39	4.08	4.13
	ULRSSM [11]	2.09	32.6	28.2	32.0	1.69	4.61	6.91
	Hybrid ULRSSM [7]	1.55	33.5	15.5	22.1	1.48	<u>3.47</u>	<u>3.95</u>
Supervised	GeoFMap [17]	2.81	30.2	25.2	24.5	2.65	4.12	4.21
	Ours (Normal)	<u>1.83</u>	2.61	4.23	4.12	<u>1.58</u>	3.11	3.38
	Ours (RGB)	-	2.57	-	-	-	-	-

propose a shear-resistant term to regularize the deformation, as described below:

$$\mathcal{L}_{shear} = \sum_i^{|\mathcal{F}|} \|\tilde{J}_i - \tilde{J}_i^{rot}\|_F, \quad (11)$$

where \tilde{J}_i^{rot} , the rotational component of \tilde{J}_i , is extracted via polar decomposition. This is inspired by the observation that nice deformation results are majorly contributed by rigid transformation and avoid shear deformation.

Optimization. The final loss is formulated as follows:

$$\begin{aligned} \mathcal{L} = & \lambda_{flow} \mathcal{L}_{flow} + \lambda_{cd} \mathcal{L}_{cd} + \lambda_{normal} \mathcal{L}_{normal} \\ & + \lambda_{identity} \mathcal{L}_{identity} + \lambda_{shear} \mathcal{L}_{shear} \end{aligned} \quad (12)$$

We iteratively optimize the per face transformation matrix \tilde{J}_i via minimizing the above loss. After the optimization converges, We use the procedure described in Equation 6 to obtain the registration results Φ_{final} . To retrieve 3d correspondences, we first find the nearest face in \mathcal{M}_{tgt} for each Φ_{final}^i . Next, Φ_{final}^i is projected onto the nearest face and the barycentric coordinate is computed to establish the final mapping between \mathcal{M}_{src} and \mathcal{M}_{tgt} .

4. Experiments

4.1. Implementation Details

To train the 2D character correspondence model, we first render 30 views each for the source and target meshes. Azimuth angles are sampled at 60-degree intervals, and elevation angles range from -30° to 50° . The resulting images are passed through DINOv2 [56] to extract 60×60 feature maps. These images are also encoded using the VAE encoder from Stable Diffusion 1.5 [63]. After adding noise

at timestep $t = 50$, the latent images are processed by the U-Net, from which features are extracted from the upsampling layers. Using a sliding window strategy, we obtain 120×120 features from the last upsampling layer. Features from earlier upsampling layers and DINOv2 are upsampled and concatenated, forming the final 120×120 feature map input to the adapter network. The weights for the geometry-grounded negative loss are set to $\lambda_{neg} = 5.0$. During inference, the adapter’s output is upsampled to 512×512 and used to compute the semantic flow map.

In the Semantic Flow Guided Registration stage, we optimize the Neural Jacobian Fields for 5,000 iterations, which takes approximately 2 minutes for meshes with 10K faces and 4 minutes for those with 40K faces. We use differentiable rendering to generate 512×512 flow and normal maps, which are used to compute flow and normal losses. The loss weights are: $\lambda_{flow} = 10.0$, $\lambda_{cd} = 1.0$, $\lambda_{normal} = 0.1$, and $\lambda_{shear} = 0.1$. The identity-preserving loss weight $\lambda_{identity}$ decays linearly from 0.01 to 0.0001 throughout the optimization.

4.2. Evaluation Metric and Dataset

We adopt geodesic error normalized by the square root of the mesh’s total surface area as the evaluation metric.

To evaluate non-isometric shape correspondence, we introduce a new benchmark dataset, Character in-the-Wild (CharW), which includes 100 meshes from various sources, such as artists and 3D generative models [45, 86]. The CharW benchmark provides 3D correspondence annotations for evaluation, with further details in supplementary material. For evaluation, we use 3DBiCar [51] and SMPL [13, 48] as source meshes, and meshes from CharW as targets to form evaluation pairs. Additionally, we evaluate on the near-isometric FAUST remesh [10], the non-isometric inter-class setting of DT4D-H [44], and CharW.

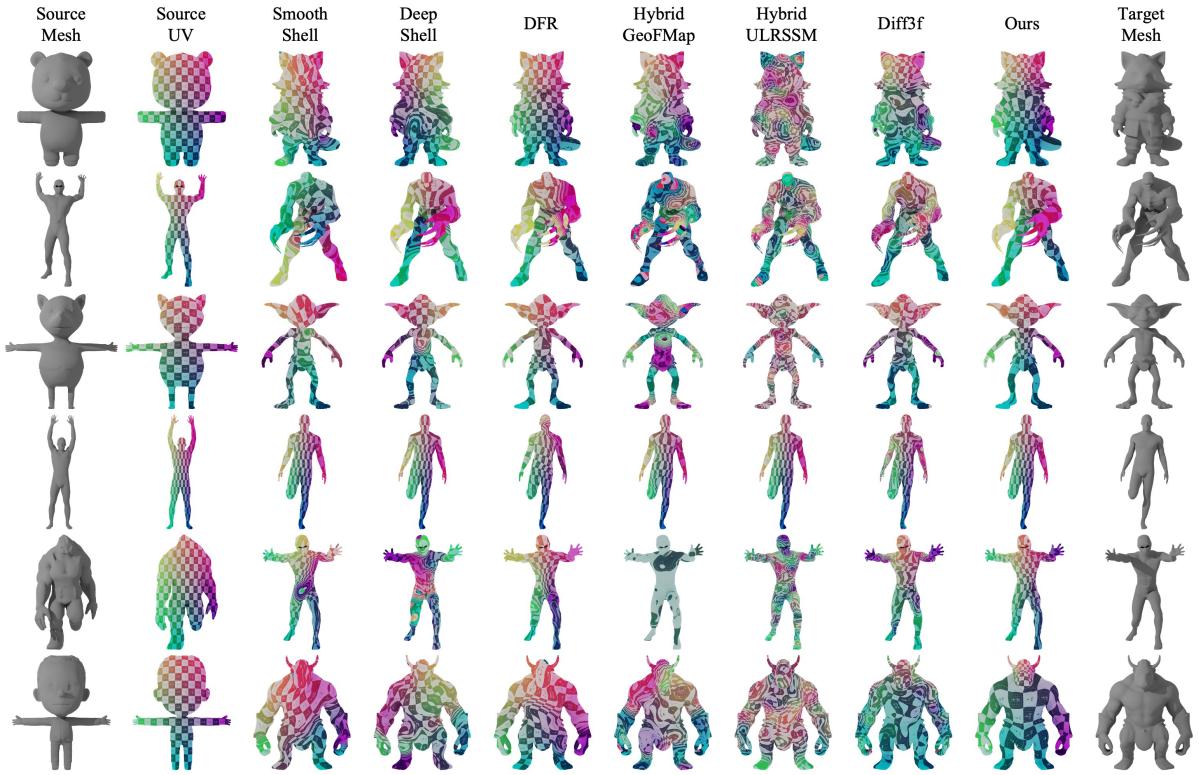


Figure 4. Visualized comparison with previous methods. All results are from the cross domain setting.

From the FAUST test set, we sample 100 mesh pairs. DT4D-H includes two variants: DT4D-H std, the standard test set from [11], and DT4D-H hard, a more challenging set that includes the Pumpkinhulk instance.

4.3. Compared with previous methods

We compare our methods with previous methods such as registration method SmoothShell [21], DeepShell [22], DFR [33], functional map series methods [7, 11, 17] and Diff3F [18]. These methods are categorized by supervision type: (1) zero-shot methods that require no training, (2) unsupervised methods trained without ground-truth correspondence, and (3) supervised methods that rely on ground-truth annotations. Our full method is categorized as a supervised method. A discussion of this design choice is provided in supplementary material. We evaluate all methods under two settings: (1) a cross-domain setting, where models are trained on 3DBiCar [51] and SURREAL [74], and tested on other benchmarks to assess generalization; and (2) an intra-domain setting, where training and testing are on the same dataset, following standard practice in prior work. We introduce the cross-domain setting since it is more practical and able to generalize to various scenario. Quantitative results are shown in Table 1, with visual comparisons in Figure 4. Our method achieves state-of-the-art perfor-

mance on non-isometric correspondence across all benchmarks. For isometric cases, we achieve the second-best result, on par with the current leading method [7]. Compared to Diff3F [18], our results highlight that leveraging foundation model features alone is insufficient. By re-purposing foundation models for 2D character correspondence and introducing Semantic Flow Guided Registration, our method significantly improves performance. Additionally, our use of RGB images as input demonstrates strong robustness and flexibility across both textured and textureless scenarios.

Table 2. Ablation study on different design choices of our method. We use RGB images as inputs when evaluating on our Character in-the-wild benchmark dataset.

Test on	Cross Domain		
	FAUST	CharW (RGB)	DT4D-H hard
baseline	2.88	3.44	6.04
Use Neural Jacobian Field	2.32	2.69	4.58
+ shear-resistant loss	2.07	2.59	4.50
+ geometry-grounded negative loss (full)	1.83	2.57	4.12

4.4. Ablation study

We first conduct ablation studies to evaluate the effectiveness of three key components: (1) the Neural Jacobian Field, (2) the shear-resistant loss, and (3) the geometry-

Table 3. Ablation study on using feature adaptor.

Test on	Cross Domain		
	FAUST	CharW (normal)	DT4D-H hard
Diff3f [18]	12.0	12.5	22.7
Ours (zero-shot)	5.60	3.48	14.1
Ours w/ feature adaptor	1.83	2.61	4.12

Table 4. Ablation study of different types of supervision given during registration.

Test on	Cross Domain		
	FAUST	CharW (RGB)	DT4D-H hard
3D correspondence supervision	2.19	2.65	4.56
Semantic flow supervision	1.83	2.57	4.12

grounded negative loss used during 2D correspondence training. Our baseline is a simple deformation model that optimizes per-vertex displacements with Laplacian smoothing, without incorporating any of the above designs. As shown in Table 2, the Neural Jacobian Field significantly outperforms the per-vertex displacement baseline. Both the shear-resistant loss and the geometry-grounded negative loss contribute to improved correspondence quality.

Furthermore, we evaluate the necessity of training a 2D correspondence model with a feature adapter. For comparison, we construct a zero-shot variant that directly uses pre-trained Stable Diffusion and DINO features for 2D correspondence. As shown in Table 3, training a feature adapter significantly boosts performance, highlighting its importance.

We also examine the impact of using semantic flow as supervision for registration. To this end, we construct a baseline inspired by Diff3F [18], where 2D features from the feature adapter network are first back-projected onto mesh vertices. Initial 3D correspondences are then obtained via nearest-neighbor search and used to supervise the registration. Results in Table 4 demonstrate that using 2D semantic flow as guidance yields superior performance.

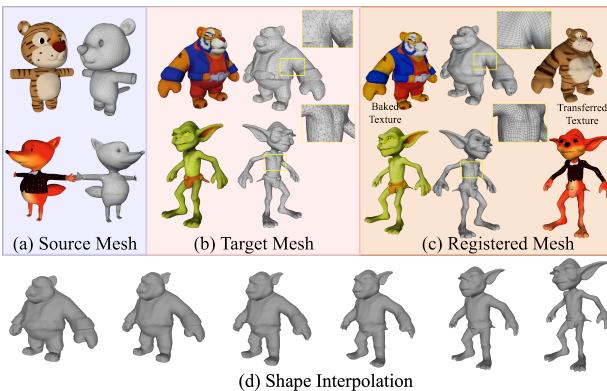


Figure 5. Application of re-topology, texture transfer and shape interpolation.

4.5. Downstream Application

In this section, we demonstrate several downstream applications of Stable-SCore, including re-topology, texture transfer, and shape interpolation using characters generated by 3D generative models [45, 86, 87]. As shown in Figure 5, we select source meshes from the 3DBiCar [51] dataset, which feature high-quality topology and artist-designed UV maps. We then apply Stable-SCore to register these source meshes to target meshes. The resulting registered meshes adopt the target’s shape while preserving the source’s topology and UV coordinates, enabling seamless texture transfer (Figure 5(c)). Moreover, by registering 3DBiCar meshes to two arbitrary target meshes, we establish dense correspondences between them and perform shape interpolation (Figure 5(d)). Additional applications—including rig transfer and animation of characters generated by text-to-3D methods—are presented in the video in the [project page](#).

5. Conclusion

We propose Stable-Score, a novel registration-based framework for stable 3D shape correspondence, addressing the challenge of non-isometric character matching. By adopting a 2D foundation model for robust 2D correspondences, our method guides a smooth registration process for accurate 3D alignment. Experiments show that Stable-Score significantly outperforms prior methods and enables downstream applications such as retopology, texture transfer, rig transfer, and shape interpolation. We also introduce the Character in-the-Wild (CharW) benchmark, a diverse dataset to further advance research in non-isometric correspondence. Our work pushes the state of 3D shape correspondence forward, opening new opportunities for both academic and practical applications.

6. Acknowledgements

The work was supported in part by the Basic Research Project No. HZQB-KCXYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, by Guangdong Provincial Outstanding Youth Fund (No. 2023B1515020055), by Shenzhen Science and Technology Program No. JCYJ20220530143604010, and No. NSFC61931024. It is also partly supported by the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, by Key Area R&D Program of Guangdong Province (Grant No. 2018B030338001), by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and by Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055).

References

- [1] Noam Aigerman, Kunal Gupta, Vladimir G Kim, Siddhartha Chaudhuri, Jun Saito, and Thibault Groueix. Neural jacobian fields: Learning intrinsic mappings of arbitrary meshes. *SIGGRAPH*, 2022. 2, 5
- [2] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007. 1, 2, 3
- [3] Shir Amir, Yossi Gandelsman, Shai Bagon, and Tali Dekel. Deep vit features as dense visual descriptors. *arXiv preprint arXiv:2112.05814*, 2(3):4, 2021. 3
- [4] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005. 13
- [5] Souhaib Attaiki and Maks Ovsjanikov. Shape non-rigid kinematics (snk): A zero-shot method for non-rigid shape matching via unsupervised functional map regularized reconstruction. *Advances in Neural Information Processing Systems*, 36:70012–70032, 2023. 3, 13
- [6] Souhaib Attaiki and Maks Ovsjanikov. Understanding and improving features learned in deep functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1326, 2023. 3, 13
- [7] Lennart Bastian, Yizheng Xie, Nassir Navab, and Zorah Lähner. Hybrid functional maps for crease-aware non-isometric shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3313–3323, 2024. 1, 2, 3, 6, 7
- [8] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part I* 9, pages 404–417. Springer, 2006. 3
- [9] Florian Bernard, Zeeshan Khan Suri, and Christian Theobalt. Mina: Convex mixed-integer programming for non-rigid shape alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13826–13835, 2020. 1, 2, 3
- [10] Federica Bogo, Javier Romero, Matthew Loper, and Michael J Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3794–3801, 2014. 6, 13
- [11] Dongliang Cao, Paul Roetzer, and Florian Bernard. Unsupervised learning of robust spectral shape matching. *ACM Trans. Graph.*, 2023. 1, 2, 3, 6, 7
- [12] Dongliang Cao, Marvin Eisenberger, Nafie El Amrani, Daniel Cremers, and Florian Bernard. Spectral meets spatial: Harmonising 3d shape matching and interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3658–3668, 2024. 1, 2, 3
- [13] Dan Casas and Marc Comino-Trinidad. Smplitex: A generative model and dataset for 3d human texture estimation from single image. *arXiv preprint arXiv:2309.01855*, 2023. 6
- [14] Seokju Cho, Sunghwan Hong, Sangryul Jeon, Yunsung Lee, Kwanghoon Sohn, and Seungryong Kim. Cats: Cost aggregation transformers for visual correspondence. *Advances in Neural Information Processing Systems*, 34:9011–9023, 2021. 3
- [15] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10286–10296, 2021. 1
- [16] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8592–8601, 2020. 1, 3, 13
- [17] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric maps: Robust feature learning for shape correspondence. *CVPR*, 2020. 1, 6, 7, 13
- [18] Niladri Shekhar Dutt, Sanjeev Muralikrishnan, and Niloy J. Mitra. Diffusion 3d features (diff3f): Decorating untextured shapes with distilled semantic features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4494–4504, 2024. 3, 6, 7, 8
- [19] Viktoria Ehm, Paul Roetzer, Marvin Eisenberger, Maolin Gao, Florian Bernard, and Daniel Cremers. Geometrically consistent partial shape matching. In *2024 International Conference on 3D Vision (3DV)*, pages 914–922. IEEE, 2024. 3
- [20] Marvin Eisenberger, Zorah Lähner, and Daniel Cremers. Divergence-free shape correspondence by deformation. In *Computer Graphics Forum*, pages 1–12. Wiley Online Library, 2019. 1, 2, 3
- [21] Marvin Eisenberger, Zorah Lahner, and Daniel Cremers. Smooth shells: Multi-scale shape registration with functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12265–12274, 2020. 2, 3, 6, 7
- [22] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. Deep shells: Unsupervised shape correspondence with optimal transport. *Advances in Neural information processing systems*, 33:10491–10502, 2020. 1, 3, 6, 7
- [23] Marvin Eisenberger, David Novotny, Gael Kerchenbaum, Patrick Labatut, Natalia Neverova, Daniel Cremers, and Andrea Vedaldi. Neuromorph: Unsupervised shape interpolation and correspondence in one go. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7473–7483, 2021. 3
- [24] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *European Conference on Computer Vision*, pages 241–258. Springer, 2025. 2, 3
- [25] William Gao, Noam Aigerman, Thibault Groueix, Vova Kim, and Rana Hanocka. Textdeformer: Geometry manipulation using text guidance. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 2, 5

- [26] Kamal Gupta, Varun Jampani, Carlos Esteves, Abhinav Shrivastava, Ameesh Makadia, Noah Snavely, and Abhishek Kar. Asic: Aligning sparse in-the-wild image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4134–4145, 2023. 3
- [27] Florine Hartwig, Josua Sassen, Omri Azencot, Martin Rumpf, and Mirela Ben-Chen. An elastic basis for spectral shape correspondence. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 1, 3
- [28] Eric Hedlin, Gopal Sharma, Shweta Mahajan, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. Unsupervised semantic correspondence using stable diffusion. *Advances in Neural Information Processing Systems*, 36, 2024. 3
- [29] Kai Hormann, Bruno Lévy, and Alla Sheffer. Mesh parameterization: Theory and practice. 2007. 1
- [30] Shuaiyi Huang, Luyu Yang, Bo He, Songyang Zhang, Xuming He, and Abhinav Shrivastava. Learning semantic correspondence with sparse annotations. In *European Conference on Computer Vision*, pages 267–284. Springer, 2022. 3
- [31] Takeo Igarashi, Tomer Moscovich, and John F Hughes. As-rigid-as-possible shape manipulation. *ACM transactions on Graphics (TOG)*, 24(3):1134–1141, 2005. 2
- [32] Allan Jabri, Andrew Owens, and Alexei Efros. Space-time correspondence as a contrastive random walk. *Advances in neural information processing systems*, 33:19545–19560, 2020. 3
- [33] Puhua Jiang, Mingze Sun, and Ruqi Huang. Non-rigid shape registration via deep functional maps prior. In *Advances in Neural Information Processing Systems*, pages 58409–58427. Curran Associates, Inc., 2023. 1, 3, 6, 7, 13
- [34] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9492–9502, 2024. 2, 3
- [35] Seungwook Kim, Juhong Min, and Minsu Cho. Transformatcher: Match-to-match attention for semantic correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8697–8707, 2022. 3
- [36] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 3
- [37] Zorah Lähner, Emanuele Rodolà, Michael M Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Alexander Dieckmann, Reinhard Klein, Y Sahillioğlu, et al. Shrec’16: Matching of deformable shapes with topological noise. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 55–60. Eurographics Association, 2016. 13
- [38] Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics*, 39(6), 2020. 5
- [39] Jae Yong Lee, Joseph DeGol, Victor Fragoso, and Sudipta N Sinha. Patchmatch-based neighborhood consensus for semantic correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13153–13163, 2021. 3
- [40] Hao Li, Robert W Sumner, and Mark Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *Computer graphics forum*, pages 1421–1430. Wiley Online Library, 2008. 1, 3
- [41] Lei Li, Nicolas Donati, and Maks Ovsjanikov. Learning multi-resolution functional maps with spectral attention for robust shape matching. *Advances in Neural Information Processing Systems*, 35:29336–29349, 2022. 1
- [42] Qinsong Li, Yueyu Guo, Xinru Liu, Ling Hu, Feifan Luo, and Shengjun Liu. Deformable shape matching with multiple complex spectral filter operator preservation. *The Visual Computer*, 40(7):4885–4898, 2024. 1, 3, 13
- [43] Yang Li and Tatsuya Harada. Non-rigid point cloud registration with neural deformation pyramid. *Advances in Neural Information Processing Systems*, 35:27757–27768, 2022. 2, 3
- [44] Yang Li, Hikari Takehara, Takafumi Taketomi, Bo Zheng, and Matthias Nießner. 4dcomplete: Non-rigid motion estimation beyond the observable surface. *IEEE International Conference on Computer Vision (ICCV)*, 2021. 6, 13
- [45] Yangguang Li, Zi-Xin Zou, Zexiang Liu, Dehu Wang, Yuan Liang, Zhipeng Yu, Xingchao Liu, Yuan-Chen Guo, Ding Liang, Wanli Ouyang, et al. Triposeg: High-fidelity 3d shape synthesis using large-scale rectified flow models. *arXiv preprint arXiv:2502.06608*, 2025. 2, 6, 8, 13
- [46] Fangfu Liu, Wenqiang Sun, Hanyang Wang, Yikai Wang, Haowen Sun, Junliang Ye, Jun Zhang, and Yueqi Duan. Reconx: Reconstruct any scene from sparse views with video diffusion model. *arXiv preprint arXiv:2408.16767*, 2024. 2, 3
- [47] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9298–9309, 2023. 2, 3
- [48] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, 2015. 4, 6, 13
- [49] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 3
- [50] Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. *Advances in Neural Information Processing Systems*, 36, 2024. 3
- [51] Zhongjin Luo, Shengcai Cai, Jinguo Dong, Ruibo Ming, Liangdong Qiu, Xiaohang Zhan, and Xiaoguang Han. Rabbit: Parametric modeling of 3d biped cartoon characters with a topological-consistent dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 4, 6, 7, 8, 13

- [52] Robin Magnet and Maks Ovsjanikov. Scalable and efficient functional map computations on dense meshes. In *Computer Graphics Forum*, pages 89–101. Wiley Online Library, 2023. 3, 13
- [53] Simone Melzi, Riccardo Marin, Emanuele Rodolà, Umberto Castellani, Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Shrec 2019: Matching humans with different connectivity. In *Eurographics Workshop on 3D Object Retrieval*, page 3. The Eurographics Association, 2019. 13
- [54] Simone Melzi, Jing Ren, Emanuele Rodolà, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. Zoomout: spectral upsampling for efficient shape correspondence. *ACM Transactions on Graphics (TOG)*, 38(6):1–14, 2019. 1
- [55] Juhong Min, Jongmin Lee, Jean Ponce, and Minsu Cho. Spair-71k: A large-scale benchmark for semantic correspondence. *arXiv preprint arXiv:1908.10543*, 2019. 4
- [56] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 3, 4, 6
- [57] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (ToG)*, 31(4):1–11, 2012. 1, 3, 13
- [58] Gianluca Paravati, Fabrizio Lamberti, Valentina Gatteschi, Claudio Demartini, and Paolo Montuschi. Point cloud-based automatic assessment of 3d computer animation courseworks. *IEEE Transactions on Learning Technologies*, 10(4):532–543, 2016. 1
- [59] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 3
- [60] Paul Roetzer and Florian Bernard. Spidermatch: 3d shape matching with global optimality and geometric consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14543–14553, 2024. 3
- [61] Paul Roetzer, Paul Swoboda, Daniel Cremers, and Florian Bernard. A scalable combinatorial solver for elastic geometrically consistent 3d shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 428–438, 2022.
- [62] Paul Roetzer, Ahmed Abbas, Dongliang Cao, Florian Bernard, and Paul Swoboda. Fast discrete optimisation for geometrically consistent 3d shape matching. *arXiv preprint arXiv:2310.08230*, 2023. 3
- [63] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021. 3, 4, 6
- [64] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 3
- [65] Jean-Michel Roufosse, Abhishek Sharma, and Maks Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1617–1627, 2019. 1
- [66] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022. 3, 13
- [67] Ruoxi Shi, Hansheng Chen, Zhuoyang Zhang, Minghua Liu, Chao Xu, Xinyue Wei, Linghao Chen, Chong Zeng, and Hao Su. Zero123++: a single image to consistent multi-view diffusion base model. *arXiv preprint arXiv:2310.15110*, 2023. 2, 3
- [68] Mingze Sun, Shiwei Mao, Puhua Jiang, Maks Ovsjanikov, and Ruqi Huang. Spatially and spectrally consistent deep functional maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14497–14507, 2023. 1, 3, 13
- [69] Mingze Sun, Chen Guo, Puhua Jiang, Shiwei Mao, Yurun Chen, and Ruqi Huang. Srif: Semantic shape registration empowered by diffusion-based image morphing and flow estimation. *arXiv preprint arXiv:2409.11682*, 2024. 3
- [70] Mingze Sun, Ruqi Huang, et al. Non-rigid shape registration via deep functional maps prior. *Advances in Neural Information Processing Systems*, 36, 2024. 1
- [71] Yang-Tian Sun, Qian-Cheng Fu, Yue-Ren Jiang, Zitao Liu, Yu-Kun Lai, Hongbo Fu, and Lin Gao. Human motion transfer with 3d constraints and detail enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4682–4693, 2022. 1
- [72] Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. *Advances in Neural Information Processing Systems*, 36:1363–1389, 2023. 3
- [73] Narek Tumanyan, Omer Bar-Tal, Shai Bagon, and Tali Dekel. Splicing vit features for semantic appearance transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10748–10757, 2022. 3
- [74] Gürkaynak Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *CVPR*, 2017. 4, 7
- [75] Vikram Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz, Dmitry Tochilkin, Christian Laforte, Robin Rombach, and Varun Jampani. Sv3d: Novel multi-view synthesis and 3d generation from a single image using latent video diffusion. In *European Conference on Computer Vision*, pages 439–457. Springer, 2025. 2, 3
- [76] Junpeng Wan, Yanxiang Bi, Zhe Zhou, and Zhou Li. Meshup: Stateless cache side-channel attack on cpu mesh. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1506–1524. IEEE, 2022. 2, 5
- [77] Duotun Wang, Hengyu Meng, Zeyu Cai, Zhijing Shao, Qianxi Liu, Lin Wang, Mingming Fan, Ying Shan, Xiaohang Zhan, and Zeyu Wang. Headevolver: Text to head avatars via

- expressive and attribute-preserving mesh deformation. In *International Conference on 3D Vision*, 2025. 2, 5
- [78] Peng Wang and Yichun Shi. Imagedream: Image-prompt multi-view diffusion for 3d generation. *arXiv preprint arXiv:2312.02201*, 2023. 2, 3
- [79] Qianqian Wang, Xiaowei Zhou, Bharath Hariharan, and Noah Snavely. Learning feature descriptors using camera pose supervision. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 757–774. Springer, 2020. 3
- [80] Xiaolong Wang, Allan Jabri, and Alexei A Efros. Learning correspondence from the cycle-consistency of time. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2566–2576, 2019. 3
- [81] Thomas Windheuser, Ulrich Schlickewei, Frank R Schmidt, and Daniel Cremers. Geometrically consistent elastic matching of 3d shapes: A linear programming solution. In *2011 International Conference on Computer Vision*, pages 2134–2141. IEEE, 2011. 3
- [82] Guangkai Xu, Yongtao Ge, Mingyu Liu, Chengxiang Fan, Kangyang Xie, Zhiyue Zhao, Hao Chen, and Chunhua Shen. Diffusion models trained with large data are transferable visual models. *arXiv preprint arXiv:2403.06090*, 2024. 2, 3
- [83] Junyi Zhang, Charles Herrmann, Junhwa Hur, Eric Chen, Varun Jampani, Deqing Sun, and Ming-Hsuan Yang. Telling left from right: Identifying geometry-aware semantic correspondence. 2023. 3, 4
- [84] Junyi Zhang, Charles Herrmann, Junhwa Hur, Luisa Polania Cabrera, Varun Jampani, Deqing Sun, and Ming-Hsuan Yang. A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence. *Advances in Neural Information Processing Systems*, 36, 2024. 3
- [85] Kaiwen Zhang, Yifan Zhou, Xudong Xu, Bo Dai, and Xinggang Pan. Diffmorpher: Unleashing the capability of diffusion models for image morphing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7912–7921, 2024. 3
- [86] Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiu, Anqi Pang, Haoran Jiang, Wei Yang, Lan Xu, and Jingyi Yu. Clay: A controllable large-scale generative model for creating high-quality 3d assets. *ACM Transactions on Graphics (TOG)*, 43(4):1–20, 2024. 2, 6, 8, 13
- [87] Zibo Zhao, Zeqiang Lai, Qingxiang Lin, Yunfei Zhao, Haolin Liu, Shuhui Yang, Yifei Feng, Mingxin Yang, Sheng Zhang, Xianghui Yang, et al. Hunyan3d 2.0: Scaling diffusion models for high resolution textured 3d assets generation. *arXiv preprint arXiv:2501.12202*, 2025. 8, 13
- [88] Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1429–1439, 2021. 1
- [89] Qi Zuo, Xiaodong Gu, Lingteng Qiu, Yuan Dong, Zhengyi Zhao, Weihao Yuan, Rui Peng, Siyu Zhu, Zilong Dong, Liefeng Bo, et al. Videomv: Consistent multi-view generation based on large video generative model. *arXiv preprint arXiv:2403.12010*, 2024. 2, 3

7. Character in-the-wild Benchmark

To better evaluate non-isometric shape correspondence, we introduce a new benchmark dataset, Character in-the-Wild (CharW). It consists of 100 character meshes collected from various sources, including artists and 3D generative models [45, 86]. Dense correspondences are annotated by manually deforming template meshes from 3DBiCar [51] and SMPL [48] to align with the collected shapes. By using a shared set of templates, ground-truth correspondences can be established between any pair of meshes in the benchmark. We compare CharW with several public character correspondence benchmarks [4, 10, 37, 44, 53] in Table 5. Unlike previous datasets, which are typically re-meshed versions of original mesh, CharW features greater diversity in shape and topology. Visualization examples are shown in Figure 6.



Figure 6. Visualized examples of the diverse shapes in our *Character in-the-wild* benchmark.

Table 5. Quantitative comparison between our *Charater in-the-wild* benchmark dataset and other character correspondence benchmark dataset.

Dataset	# of identities	shape variance	Non-isometric	various sources
FAUST [10]	10	small		
SCAPE [4]	1	small		
SHREC'19 [53]	44	small		
TOPKIDS [37]	1	small		
DT4D-H [44]	10	large	✓	
CharW(Ours)	100	large	✓	✓

7.1. CharW Dataset Curation Details

Data Collection We collect meshes from two main sources: generated meshes and artist-crafted meshes. The generated meshes come from Tripo [45] and Rodin [86], while the artist-designed meshes are sourced from Mixamo and Sketchfab. The dataset includes 38 meshes from Mixamo, 23 from Rodin, 22 from Tripo3D, and 40 from Sketchfab. Several criteria were followed during data collection: 1. All meshes represent bipedal characters, including both human and humanoid figures. 2. The meshes must be of high quality, free of significant distortion or artifacts, and avoid low-poly models. 3. We selected characters with diverse body types such as fat, slim, large-headed, and small-headed, to ensure a variety of non-isometric shapes.

Correspondence Annotation CharW benchmark provides correspondence annotations for evaluation. The annotation process begins by selecting the SMPL neutral template mesh and three meshes from 3DBiCar [51] as deformation templates. For each target mesh, a professional artist first selects the template that most closely resembles the target. The artist then annotates key-point correspondences between the source and target meshes, ensuring that the key points are semantically aligned. On average, 60 to 80 key points are annotated per mesh. After annotation, the artist uses ZBrush’s warp add-on to align the source mesh with the target. The annotated key points are refined until the results are satisfactory. If needed, the artist manually adjusts the warped mesh using sculpting tools. Finally, the warped mesh is processed with Blender’s Shrinkwrap modifier to precisely match the target.

8. Supplementary

8.1. Limitation and Future work

There are several limitations stemming from the 2D correspondence model: (1) significant initial rotation misalignment, which leads to incorrect 2D correspondences and distorts the deformation process; (2) severe occlusion, resulting in missing 2D correspondences; and (3) difficulty handling complex structures due to the low resolution of 2D features, making it challenging to capture details such as fingers, as shown in the first row of Fig 9.

The deformation process also has limitations: (1) topological noise, such as the body and arms merging together, hinders deformation; (2) it inherits limitations from NJF, including the inability to handle partial shapes. An interesting direction for future work is to move beyond the limitations of low-resolution 2D prior models by integrating native 3D prior models, such as large 3D generative models [45, 86, 87], which could potentially address these issues.

8.2. Discussion on Supervision Types

Previous functional map-based methods [5, 6, 16, 17, 33, 42, 52, 57, 68] are mostly unsupervised. Supervised methods [16] are rare in shape correspondence, and previous supervised attempts have underperformed compared to unsupervised methods. While supervised methods generally outperform unsupervised ones in most computer vision tasks, this is not the case for shape correspondence. We hypothesize that design flaws in earlier supervised methods, such as strong reliance on LBO’s basis or DiffusionNet’s prior [66], lead to unsatisfactory results and overfitting problems, suggesting significant room for improvement.

To address this, we propose a supervised, registration-based method that avoids reliance on functional maps

or DiffusionNet. Our approach achieves state-of-the-art performance on various non-isometric benchmarks. Although it requires ground-truth (GT) supervision, the annotation cost is relatively low, requiring only sparse 2D/3D key points or dense correspondences via template warping (same as the CharW annotation process). This makes our method highly cost-effective, therefore, the need for GT supervision is no longer a limitation compared to unsupervised approaches.

8.3. Beyond Characters

We also test the applicability of our method to other domains, such as animal shape correspondence, which is commonly evaluated in non-isometric shape matching. Additionally, we compare our method on the SMAL dataset, as shown in Table 6 and Figure 7. The quantitative comparisons are conducted under two setups: training on a character dataset or on the SMAL dataset, with testing performed on the SMAL dataset. “Ours (zero-shot)” refers to a zero-shot version of our method, where feature adapters are removed, and no further fine-tuning is required. Our method outperforms others and demonstrates its ability to generalize to domains where ground truth correspondences are available for training. Furthermore, the zero-shot setup of our method is versatile, achieving strong performance across various tasks.

Table 6. Quantitative comparison between our methods and previous methods on SMAL dataset.

Test on SMAL	Train on character	Train on SMAL
ULRSSM	28.5	3.63
Hybrid ULRSSM	44.0	3.11
Ours (zero-shot)	8.91	8.91
Ours (full)	17.01	2.65

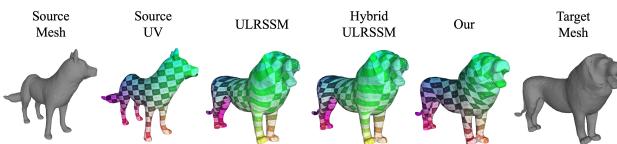


Figure 7. Comparison on SMAL, all methods are trained on SMAL.

9. More results

More results on the DT4D dataset and CharW dataset are shown in Figure 8, 9, 10 and 11. It can be observed that our Stable-Score achieves precise registration while simultaneously preserving the source topology, thereby demonstrating significant potential for re-topology applications.

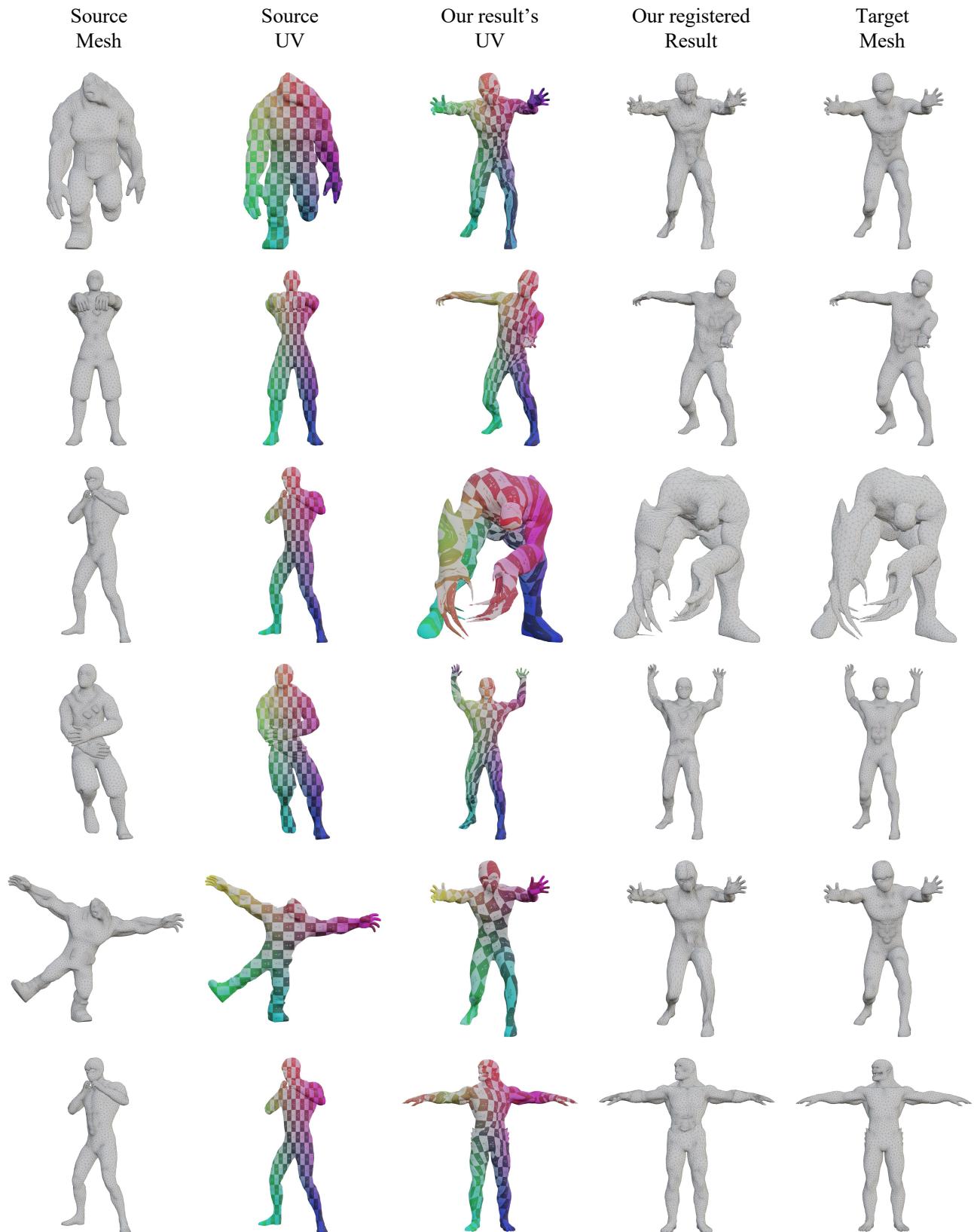


Figure 8. Visualized results produced by Stable-Score on DT4D dataset.

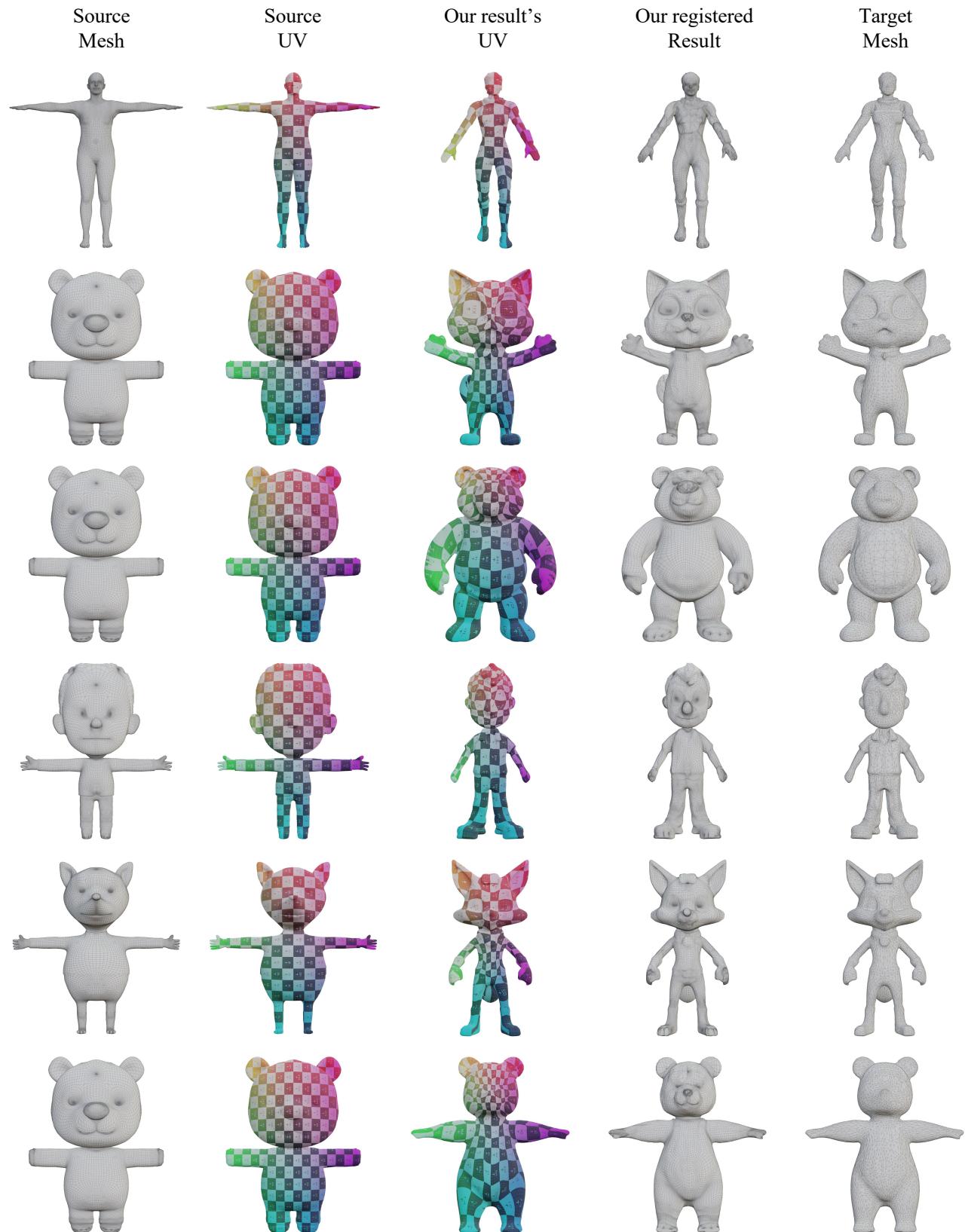


Figure 9. Visualized results produced by Stable-Score on our CharW dataset.

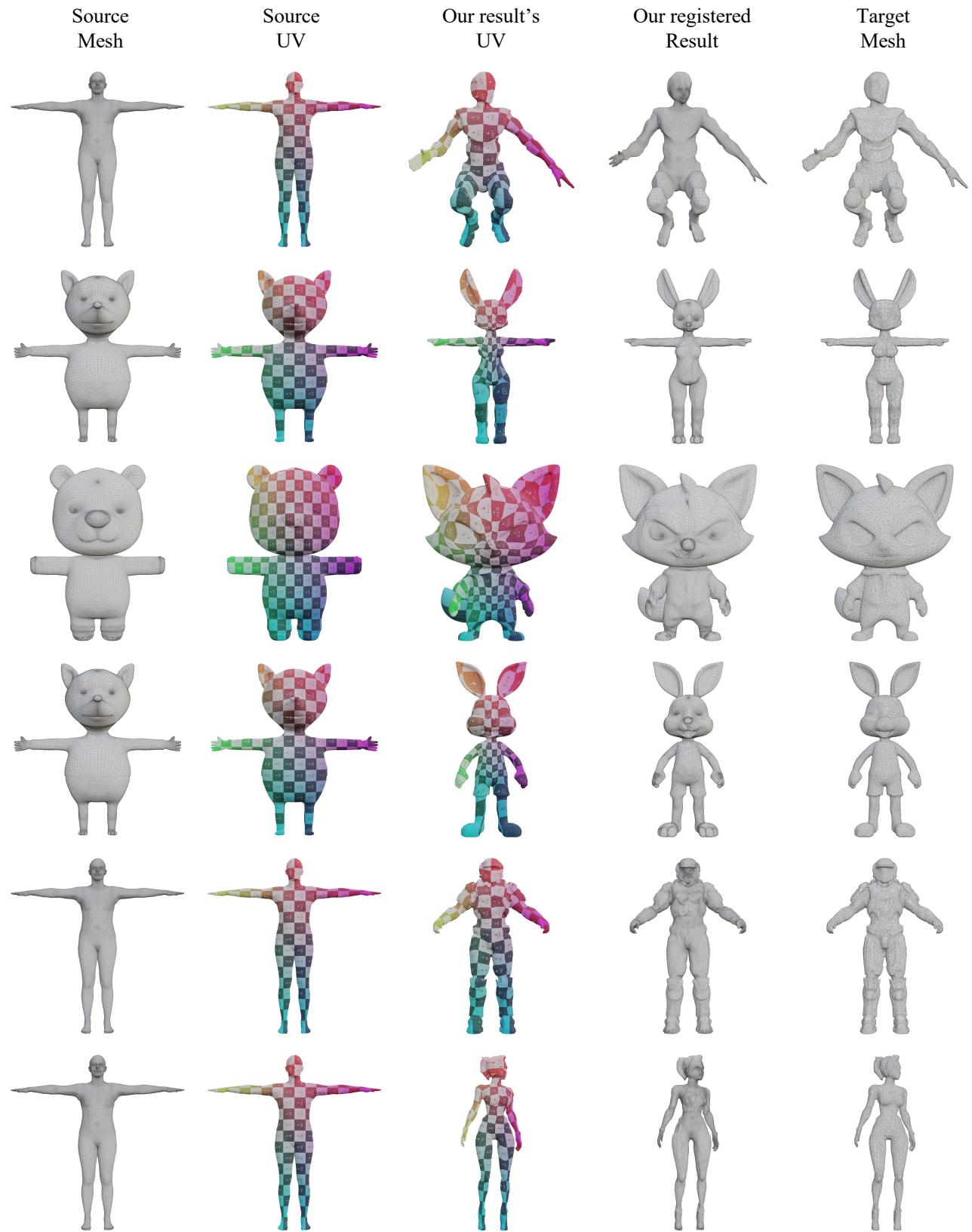


Figure 10. Visualized results produced by Stable-Score on our CharW dataset.

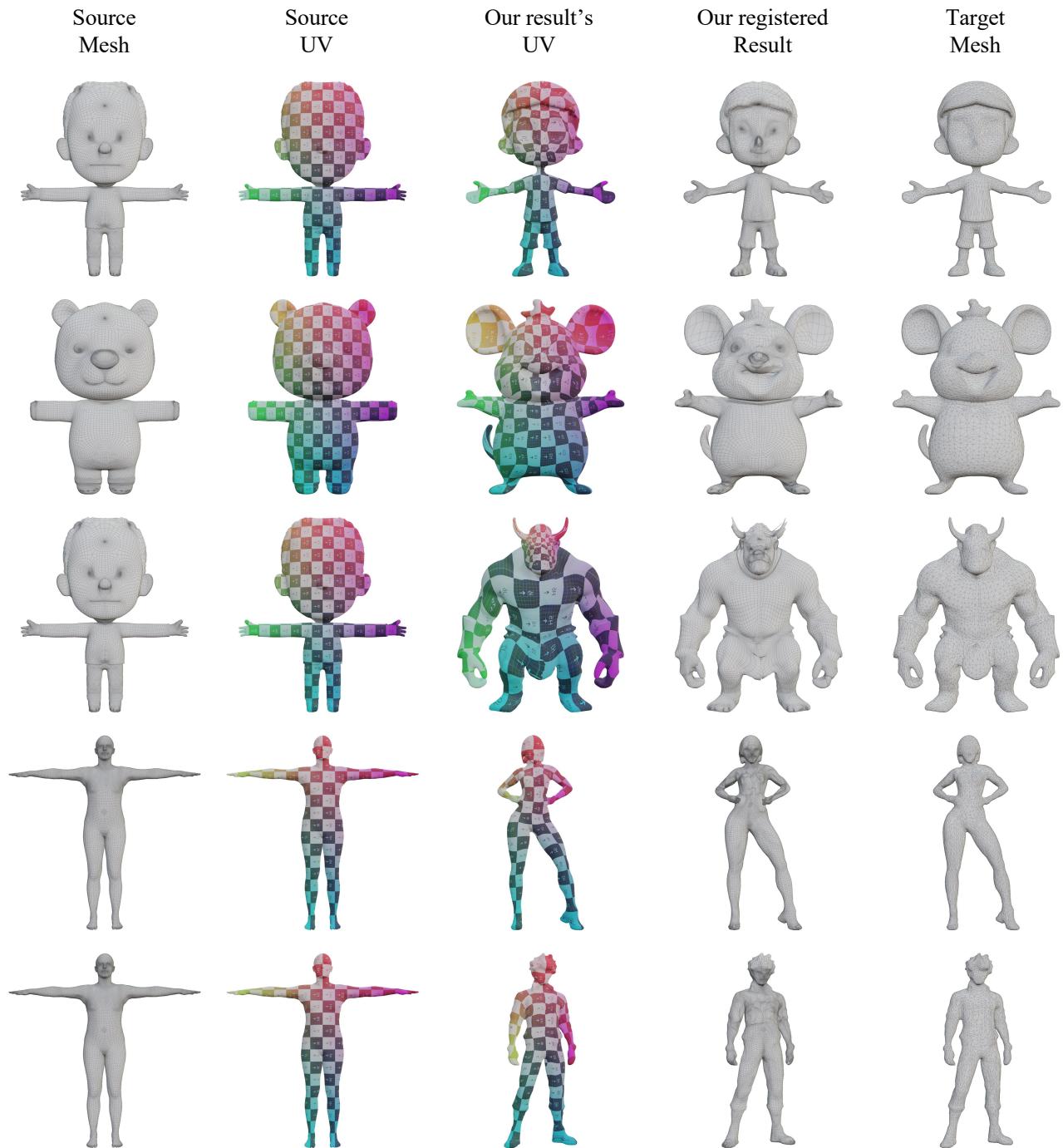


Figure 11. Visualized results produced by Stable-Score on our CharW dataset.