Bas-relief modelling from enriched detail and geometry with deep normal transfer

Meili Wang [a,*,1,2], Li Wang [b,1,3], Tao Jiang [b], Nan Xiang [b], Juncong Lin [e], Mingqiang Wei [c], Xiaosong Yang [b], Taku Komura [d], Jianjun Zhang [b]

[a] *Northwest A&F University, Yangling, China*
[b] *Bournemouth University, UK*
[c] *Nanjing University of Aeronautics and Astronautics, China*
[d] *University of Edinburgh, UK*
[e] *Xiamen University, China*

ARTICLE INFO

ABSTRACT

Detail-and-geometry richness is essential to bas-relief modelling. However, existing image-based and model-based bas-relief modelling techniques commonly suffer from detail monotony or geometry loss. In this paper, we introduce a new bas-relief modelling framework for detail abundance with visual attention based mask generation and geometry preservation, which benefits from our two key contributions. For detail richness, we propose a novel semantic neural network of normal transfer to enrich the texture styles on bas-reliefs. For geometry preservation, we introduce a normal decomposition scheme based on Domain Transfer Recursive Filter (DTRF). Experimental results demonstrate that our approach is advantageous on producing bas-relief modellings with both fine details and geometry preservation.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Bas-relief, a special type of sculpture that figures are slightly emerged from a background, is a bridge between 2D drawings and 3D sculptures. Bas-relief has received considerable attentions in recent years since it can be viewed from many different angles without causing distortion of the figures. Due to this desirable intrinsic nature, bas-reliefs as an art form has been very popular since prehistorical time [1]. They are now treated as either a single piece of artwork or decorations for walls, monuments, furniture, medals, potteries etc. In recent years, more and more researchers in Computer Graphics have developed approaches to fulfill the stylistic design purpose of bas-reliefs. Fig. 1 shows a couple of real

bas-relief examples designed by artists and digital ones generated by the framework described in this paper. Please note the fine details of the bas-relief examples in Fig. 1.

Detail and geometry richness is an essential key to artistic creations. Missing any of them, the generated bas-relief will be impossible to convey the ideas from the artists. For example, the relief in the right of Fig. 1, the scaly textures and petal details present realistic fish and lotus flowers. The display of complex geometry shapes and capture of such realistic fish in motion make them stand out artistically.

In general, digital bas-reliefs come from two types of sources: 2D images and 3D models. Methods based on natural images have ill-posed problems ([2]3) in nature, and the approaches based on 3D models mainly focus on designing sophisticated non-linear depth compression algorithms which typically incur a high computational cost. Some recent works (e.g.,[4–6]) attempt to simplify the bas-relief modelling problem by working on the normal images which contain both pixel-level detailed appearance (2D information) and the normal information of the geometry leading to stereoscopic perception (3D look). The bas-reliefs are modelled in normal image space rather than in object space, which solves the ill-posed problems that other image-based

---

* Corresponding author.
  E-mail address: wml@nwsuaf.edu.cn (M. Wang).
[1] Equal contribution.
[2] MeiliWang is with the College of Information Engineering, Northwest A&F University, China. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China. Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling 712100, China.
[3] Li Wang is with the NCCA, Bournemouth University, UK.
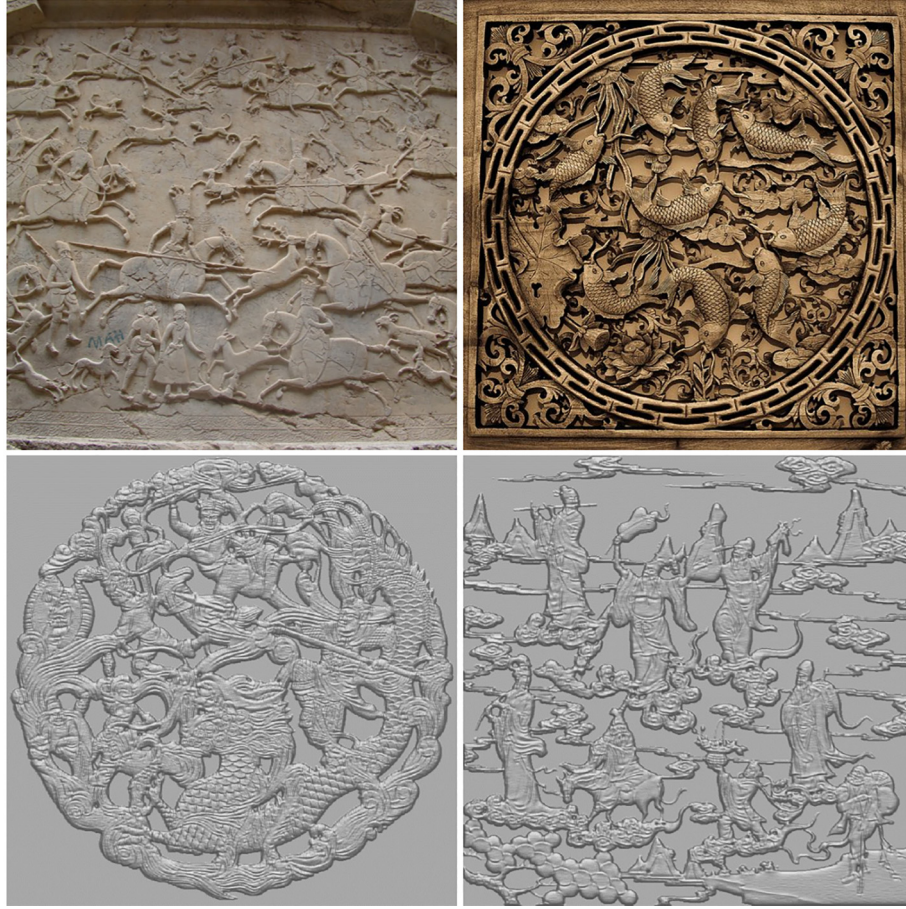
M. Wang, L. Wang, T. Jiang et al.

Fig. 1. The upper row shows artistic bas-relief works. The lower row shows our digital bas-relief results. We recommend readers to view the electronic version for details.

approaches suffer and removes the need of sophisticated depth compression algorithms.

Realistic bas-reliefs should present both detailed appearances and stereoscopic perception. For detail transfer, existing works ([4,5]) tend to rely on straightforward image processing techniques like cut-and-paste and decompose-and-compose, which often lead to imperfect composition results. For example, the cut-and-paste operation cannot preserve the normals of the target. The decompose-and-compose [7,8] operation requires larger detail patches than target patches while it cannot manipulate target surfaces or cause scaling issue for textures. For geometry preservation, existing methods (e.g., [6]) decompose the normal field into a base layer and a detail layer by directly subtracting the base normal value from the original normal, which causes unexpected triangle distortions in the resultant bas-reliefs.

In this paper, we present a novel bas-relief modelling method which is able to overcome the above-mentioned issues, and produces bas-reliefs with rich details as well as preserving the geometry intact. Inspired by image style transfer [9], we propose a semantic neural network of normal transfer which treats detail transfer as a style transfer problem. The proposed network is capable of transferring the fine details from a source normal image to the target normals in arbitrary shapes and scales. In addition, to generate transferred details on desired areas of the normal images, we adopt a visual attention mechanism [10,11] and use object parsing [12] to predict the corresponding masks. We extract rich texture areas and desired target areas from input normals into masks, then these masks are used in the proposed

semantic normal transfer network to produce the transferred normals. For geometry preservation, we present a normal decomposition scheme based on Domain Transfer Recursive Filter (DTRF) to enhance the geometry properties. We adopt the local shaping and global blending steps from [6] to construct the mesh of a bas-relief from detail and structure layers obtained by normal decomposition.

The overview of our framework is shown in Fig. 2. Our digital bas-relief modelling pipeline which takes 3D models as inputs, manipulates normal fields and generates diverse visual effects of texture transfer, structure and detail preservation. The proposed semantic neural network of normal transfer facilitates the design and texture transfer of bas-reliefs and makes the generation process in an efficient and intelligent way. The main contributions of this paper are:

• **Semantic neural network of normal transfer**. Our proposed semantic neural network of normal transfer learns the texture and structure representations, and then recombines them to generate a new normal image which shows the similar texture patterns of the source normal image and similiar structure surface of the target normal image. It is capable of taking arbitrary sizes and shapes of normal images as inputs, and then synthesizing them into a new texture.

• **Geometry preservation without introducing artefacts.** By considering the orientation of vector rotations, our normal decomposition scheme obtains a structure layer and a detail layer with continuous and more natural edges and shapes, which contribute to produce artefacts free bas-relief modellings.
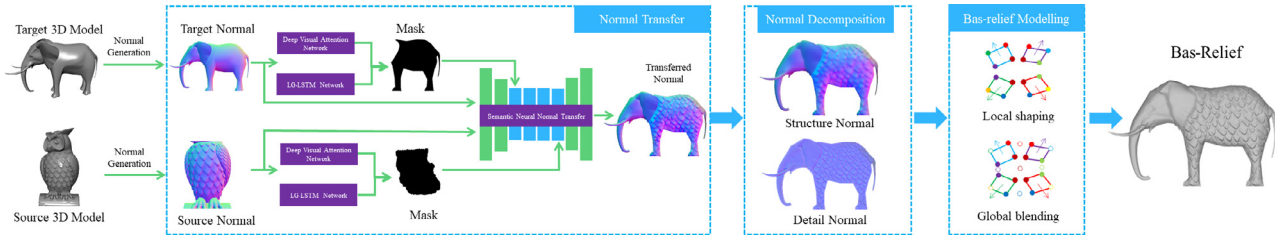
Fig. 2. The overview of the proposed approach. Our approach contains three stages: normal transfer, normal decomposition and bas-relief modelling. Normal transfer completes the task that transfers the fine details from source normal to target normal; Normal decomposition creates structure normal and detail normal with enhanced geometry properties; Bas-relief modelling constructs bas-reliefs from structure and detail normal obtained from normal decomposition.

2. Related work

In the last two decades, generating digital bas-reliefs from 3D scenes or 2D images has been a thriving subject in computer graphics. A detailed review work can be found in [13,14], which classify methods into direct modelling, image-based and shape-based modelling. Direct modelling involves experts' laborious work. Image-based 3D construction inherently has the ill-posed problem. Most 3D model-based works create bas-relief by either directly compressing the depth or by working in the gradient domain, where the final model is obtained by solving a Poisson equation. Most cases aim to generate a reproducible and manipulable mesh, which can later be used and enhanced with some more advanced graphical tools. As a result, specific colour and texture of the relief are usually not considered in these previous research. Our work is designed to solve this problem by using the learning scheme to transfer textures from 3D source models to 3D target models.

The following review work focuses on bas-relief modelling, style transfer and learning techniques in recent years.

2.1. Bas-relief modelling

2.1.1. 3D model-based methods

By applying feasible constraints on a given 3D shape or scene, Christian et al. [15] propose a novel view-dependent surface representation which allows us to cast the optimization as a quadratic program. Ji et al. [16] propose a highly efficient two-scale bas-relief modelling method on GPU, in which the input 3D scene is first rendered into two textures with depth information and normal information respectively. The depth map is then compressed to produce a base surface with level-of-depth, and the normal map is used to extract local details. Finally, the local feature details are added back to the base surface to produce the final result. Based on 3D models, Zhang et al. [17] propose a series of gradient-based algorithms which operates directly on a triangular mesh and ensures that the mesh topology remains unchanged during geometric processing. They also present two types of shape editing tools that allow the user to interactively modify the bas-relief and exhibit a desired shape. Given target shapes, viewpoints and space restrictions, Zhang et al. [18] find a global optimal surface that delivers the desired appearance when observed from the designated viewpoints, which could guarantee exact depth bounds of per-vertex. Zhang et al. [19] treat an input object as a continuous relief depth map and use mesh intersection to paste the relief on the target object based on empirical mode decomposition in multi-scale levels [20].

In contrast to model-based methods, we start from image normal operation to make the process conciser and more efficient.

2.1.2. Image-based methods

Generating bas-relief from natural images and photographs are intuitive. However, we all know that there is an ill-posed problem to recover 3D shape from a single image, since colour, luminance and texture in an image could not reflect the geometric attributes of the objects properly, especially for objects with complex materials. To overcome this problem, some researches are restricted to some special types of bas-relief from certain images. For example, Wu et al.[21], Wu et al. [22], and B. Sohn [23] concentrate on bas-relief modelling from human face photographs. Zhang et al.[19] pay special attention to model Chinese calligraphy reliefs. Li et al. [24] aim at restoring brick and stone alike relief from single rubbing image in a visually plausible manner. Zhang et al. [25] concentrate on portrait relief modeling. Some researches are based on Shape From Shading (SFS) which requires human interaction [2,3,26].

Unlike generating bas-relief from natural images, some recent works start from normal images [27]. Ji et al. [4] present a novel framework to design bas-relief in normal image space instead of object space which is capable of producing different styles of bas-relief and allows intuitive style control. Their method generates high-quality bas-relief which enables a variety of applications, such as the cut-and-paste operation and bas-relief modelling on curved surface. Recently, Ji et al. [5] extend their previous work with a layer-based editing approach for normal images to generate more diversified styles of results, and is capable of transferring details from one region to another. Similar to Ji et al.'s work [16], Wei et al. [6] decompose image normal of an input 3D model into a smooth base layer and a detail layer in order to contribute to both features of structure-preserving and detail-preserving.

Inspired by these works, our method starts from normal images of 3D meshes, using detail transfer to achieve the detail-enriched relief model with a relatively conciser operation.

2.2. Detail transfer

With the growing availability of abundant 3D mesh collections, some research works attempt to transfer textures [28] or details to geometric shapes.

Mitra et al. [29] propose an unsupervised learning method to transfer texture information from images of real objects to 3D models of similar objects by tackling the reconstruction problem of a set of base texture. Huang et al. [30] present a novel user-assisted approach to extract a non-parametric appearance model from a single photograph of a reference object (whose geometric structure roughly approximates that of the target object). A novel alignment algorithm is proposed to enable accurate joint recovery of the geometric detail and reflection. Berkiten et al. [31] propose a method which transfers details (specifically, displacement maps) from existing high-quality 3D models to simple shapes. They adopt metric learning to find a combination of geometric features that successfully predict detail-map similarities on the source mesh; then they use the learned feature combination to drive the detail transfer in texture space.

Compared to those mentioned detail transfer methods, our work aims to transfer high quality details to arbitrary shapes and preserves those details for the purpose of bas-relief generation.

2.3. Image style transfer

Recently, Convolutional Neural Network (CNN) shows its power as a rich feature provider in visual perception, which boosts several computer vision research areas such as image classification [32] and object detection [33]. Inspired by the success of visual perception using CNN, Gatys et al. [9] initialize an artificial algorithm for the creation of artistic images, which possesses the capability of learning and recombining feature representations of content and style images. Feeding images into the pre-trained CNN, they presented the semantic structure representations of a photo as neural responses of CNN, and referred to the artistic style of an artwork as its texture representations via the spatial summary statistic of neural responses. These high-level representations actually store the learned distribution of structure details in the content image and texture patterns in the style image. By manipulating these high-level representations inside CNN, they are able to generate a new image which preserves the spatial structure details of a photo and artistic style of an artwork image.

3. Semantic neural normal transfer

A point in a normal image indicates a normal vector, thus spatial structures and texture patterns are actually the orientation differences between normal vectors. As far as we know, deep neural networks [34,32] are capable of learning these structures and patterns from images, which indicates that the orientation differences of normal vectors can be captured and stored as neural responses in networks. Based on this, we propose a semantic deep neural network of normal transfer which accomplishes the detail transfer task by learning and recombining spatial structures and texture patterns from input normal images. In addition, we adopt Deep Visual Attention Network (DVA) [10] and LG-LSTM Network [12] to generate our mask images.

Fig. 3 illustrates the basic procedure of extracting areas for masks. The DVA network produces human eye attention areas for the input normal map, and LG-LSTM network generates parsing segmentations based on object parts (e.g., head, body and legs etc). Since the attention areas usually lie on heads or faces [10] which are identity areas with rich details, we utilize this feature to extract our masks. For example, DVA network produces highlight area in the elephant normal on its head, while LG-LSTM network generates part segmentations. To target normals, we always want areas with less details in order to transfer rich details into these areas, thus we choose segmented areas without attention areas (i.e., regions exclude head in Fig. 3) which consist of final target mask areas. To source normals, in contrast, we always want areas with rich details, thus we choose segmented areas where highlight areas lie.

Fig. 4 shows the architecture of the proposed semantic neural network of normal transfer. In this work, the source normal image provides the texture patterns, the target normal image provides the spatial structures and the pixel-level binary mask images indicate the regions that are valid for texture transfer. At the beginning of the optimization process, the target normal image and its corresponding mask, the source normal image and its corresponding mask are passed into the network and their features are learnt in the network. Then our network starts from the target normal image and gradually syntheses it into a new normal result via optimization iterations. This optimization process minimizes the Euclidean distance between texture and structure representations. For the given source normal image x_{tex}, target normal image x_{str} and masks m_{tex} and m_{str}, the proposed network searches a new stylized normal image $\widehat{x}$ by minimizing the following loss term:

$$\mathscr{L}(\widehat{x}, x_{str}, x_{tex}, m_{str}, m_{tex}) = \alpha \mathscr{L}_{str}(\widehat{x}, x_{str}, m_{str}) \\ + \beta \mathscr{L}_{tex}(\widehat{x}, \kappa, x_{tex}, m_{tex}) + \delta \mathscr{R}_{tv} \tag{1}$$

with:

$$\mathscr{R}_{tv} = \sum_{i,j} \left(\left(\widehat{x}_{(i,j+1)} - \widehat{x}_{(i,j)} \right)^2 + \left(\widehat{x}_{(i+1,j)} - \widehat{x}_{(i,j)} \right)^2 \right) \tag{2}$$

where the **structure loss** $\mathscr{L}_{str}$ penalizes the difference of valid structure representations between x_{str} and $\widehat{x}$; the **texture loss** $\mathscr{L}_{tex}$ penalizes the difference of valid texture representations between x_{tex} and $\widehat{x}$. To encourage the spatial smoothness in the generated image $\widehat{x}$, we also add a **total variation regularization** in our network. α and β denote the weights to balance the structure component and texture component of the stylized result $\widehat{x}$ while κ and δ respectively denote the weights of texture scales and smoothness.

Let the matrix $\phi_j(\cdot) \in \mathbb{R}^{N_j \times M_j}$ denotes the vectorized feature maps representing the neural responses in a layer j where N_j is the number of channels and M_j is $Height \times Width$ of the corresponding feature maps, and the mask m_{str} indicates the valid regions for structure preservation. The structure loss $\mathscr{L}_{str}$ is defined as the mean square error between the two valid feature representations $\phi_j(x_{str})$ and $\phi_j(\widehat{x})$ in the masked area:

$$\mathscr{L}_{str}(\widehat{x}, x_{str}, m_{str}) = \sum_{j \in J_s} m_{str} \cdot \left(\phi_j(x_{str}) - \phi_j(\widehat{x}) \right)^2 \tag{3}$$

where J_s denotes the set of layers in a pre-trained VGG-19 network [35] in which the structure loss is computed. Gatys et al. [36] have discovered that the Gram-based correlations of neural responses can be exploited as the texture representations. Hence, the texture loss $\mathscr{L}_{tex}$ is denoted as the squared Euclidean distance between the scaled texture normal representation $\kappa \cdot x_{tex}$ and the generated new normal representation $\widehat{x}$:

$$\mathscr{L}_{tex}(\widehat{x}, \kappa, x_{tex}, m_{tex}) = \sum_{j \in J_t} \kappa \cdot m_{tex} \\ \cdot \left(\psi(\phi_j(\kappa \cdot x_{tex})) - \psi(\phi_j(\widehat{x})) \right)^2 \tag{4}$$

where J_t denotes the set of layers in the pre-trained VGG network in which the texture loss is computed, $\psi(\phi_j(\cdot)) = \phi_j(\cdot) \cdot \phi_j(\cdot)^T \in \mathbb{R}^{N_j \times N_j}$ is the Gramian Matrix [36], which is used to represent the texture information.

4. Image-based normal decomposition

Normal decomposition aims to extract a structure layer L_s and a detail layer L_d from the original normal field L_o. Generally, the structure layer L_s is achieved by applying a normal filtering to smooth L_o, and L_d is obtained by subtracting L_s from L_o [6]. However, the detail layer actually presents the orientation differences between normal vectors in structure L_s and L_o. Thus we adopt the idea [5] which takes the orientation into account, and we propose a normal decomposition scheme based on an edge-preserving technique Domain Transfer Recursive Filter [37]. Let $P_{(i,j)}$ and $P_{(i,j-1)}$ denote two adjacent points in a normal image I, the result $J(P_{(i,j)})$ of DTRF in [37] is defined as:

$$J(P_{(i,j)}) = \frac{1}{K_p} \sum_{k=1}^{c} (1 - a^d) \cdot I^k(P_{(i,j)}) + a^d \cdot J^k(P_{(i,j-1)}) \tag{5}$$

where $a = exp\left(-\sqrt{2}/\sigma_s\right)$ and $d = 1 + \frac{\sigma_s}{\sigma_r}|I(P_{(i,j)}) - I(P_{(i,j-1)})|$. σ_s and σ_r are respectively standard deviation and deviation range. K_p denotes the scaling factor that normalizes $J(P_{(i,j)})$ to a unit vector and c denotes the channel index (e.g., RGB). To achieve a symmetric response of Eq. (5), the DTRF filter is applied twice: for a normal
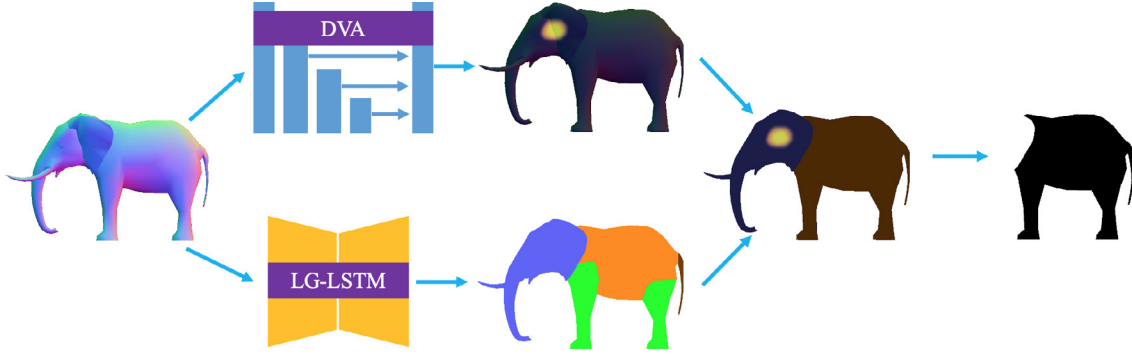
Fig. 3. The diagram of mask area extraction. The DVA network generates human eye attention area on elephant's head which is an identity region with richer details than its other parts, and the LG-LSTM network segments the elephant into various regions based on object parts. Then we choose regions exclude head as our mask. In general, we choose areas without eye attention for target normals but with highlight regions for source normals. To target normals, we always want areas with less details in order to transfer rich textures into these areas. To source normals, however, areas with highlight attention are more desired.
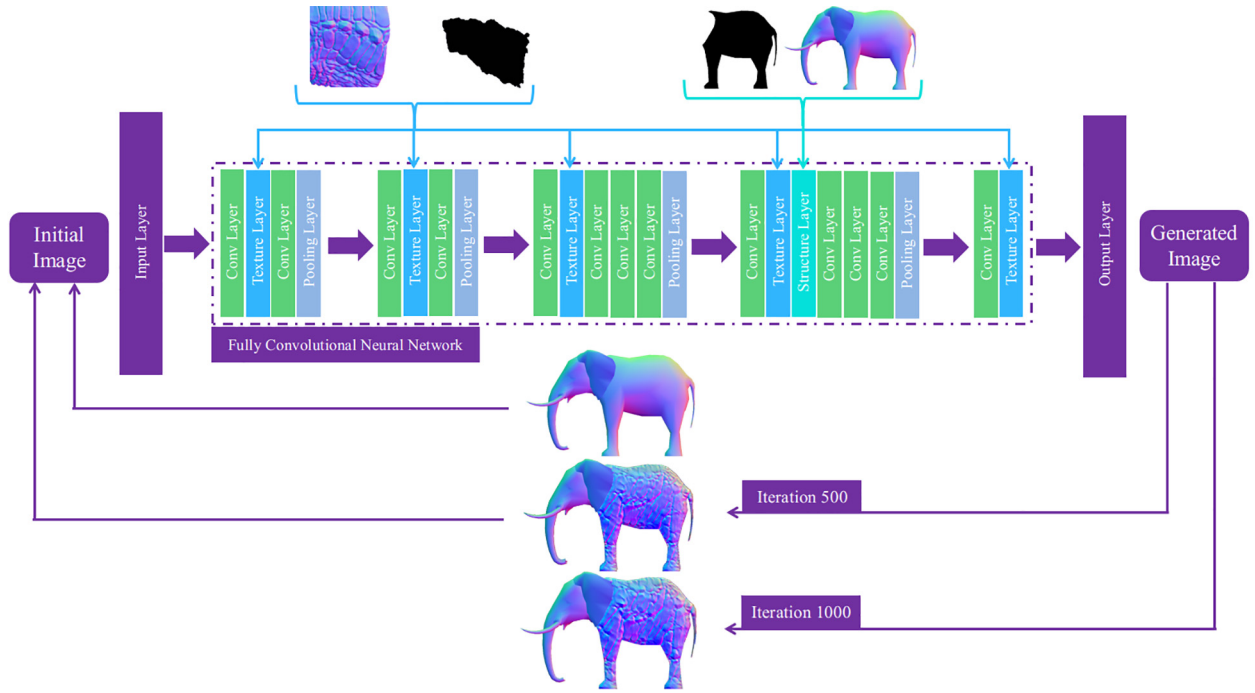
Fig. 4. The overview of our proposed semantic neural normal transfer network. The network takes source/target normal images and their corresponding masks as inputs, then computes a new generated normal image with transferred details from source normal via an optimization process.

image *I*, Eq. (5) is performed left-to-right (top-to-bottom) and then right-to-left (bottom-to-top). To simplify the user-specified parameters, we set σ_r to 1.0 which works well in all of our experiments.

Fig. 5 shows one example using our normal decomposition. We define our normal decomposition filter based on an edge-preserving technique Domain Transform Recursive Filter (DTRF) [37] which is capable of working on colour images at arbitrary scales without the need of resorting to subsampling or quantization. We now give the reason why DTRF is better than BF [38] for normal smoothing. To DTRF, the output of Eq. (5) tends to be $I(P_{(i,j)})$ itself as increased d leading a^d to zero, thus the value of a point on edge is preserved well and barely affected by its neighbouring non-edge point. For example, non-edge point $P_{(i,j-1)}$ (white color) in (b) of Fig. 6 affects edge point $P_{(i,j)}$ much less than point $P_{(i,j+1)}$ (Eq. (5) is applied twice) as its weight a^d of $P_{(i,j-1)}$ tends to be zero. In contrast, values of points on edges computed by BF can still be affected by non-edge points (i.e., points in white color

in 5×5 window size (c)), and this influence could be enhanced after a few iterations, especially when the neighbouring points of an edge point share similar values. For example, in Fig. 7, the edge of top-right corner in the zoom-in of structure normal (BF) is almost wiped out and updated into the smiliar color of non-edge neighbouring points.

Detail enhancement. To demonstrate the effectiveness compared to value subtraction, we compare the detail enhancement results obtained by Wei et al. [6] and our normal decomposition in Fig. 8. To achieve detail enhancement effect, we simply enhance the detail layer via increasing brightness and contrast in Adobe Photoshop CS6. Note that our detail layer looks much more natural than the result of Wei et al. [6] in structure surface and boundaries between edge and surfaces.

Now we analyse the reason why our normal decomposition achieves more natural bas-relief modellings than Wei et al. [6]. We consider the orientation for the normal subtraction. Each point
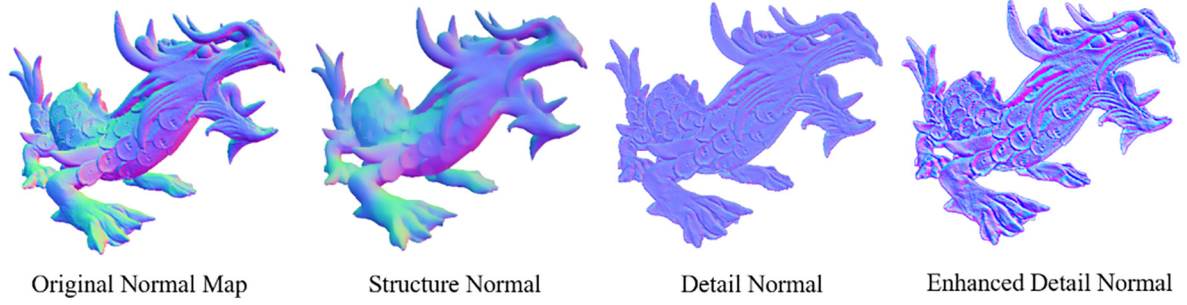
Fig. 5. One example of our normal decomposition results $\{\sigma_s = 10, \sigma_r = 1.0\}$ and we enhance the detail normal in Photoshop.
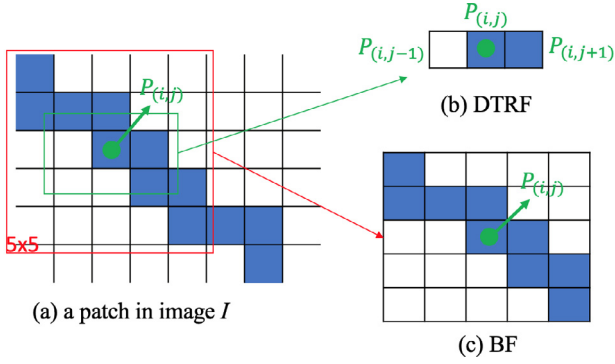

Fig. 6. Demonstration of DTRF and BF in computing values of edge points. The blue points are assumed as edge points while other points are non-edge points. The point $P_{(i,j)}$ is the current point computed by Eq. (5).

in the normal image indicates a normal vector. Thus the normal differences between two points in base and original normal are actually orientation differences which should follow the vector rotation. As the orientations of the corresponding points in base and original normal field are continuous, the orientations of new vectors computed have the property of continuity as well. Therefore, the detail layer containing those orientation differences gives continuous and reasonable edges and surfaces in image space. Wei et al. [6] perform normal decomposition on mesh level via a vector length threshold θ which follows a certain Gaussian distribution, then the small triangles with points having shorter vector length are remained in the detail layer (c.f. the orange rectangle shown in the bottom-middle of Fig. 8). However, the orientations between remained neighbour triangles are not continuous which leads to unexpected triangle distortions and damages the structure surface in bas-relief modelling results when detail enhancement is applied (c.f. the red and green rectangles in Fig. 8). Our normal decomposition has no such damage.

5. Bas-relief modelling

In this section, we construct the bas-relief model based on the decomposed structure normal map L_s and detail normal map L_d. To generate the surface from decomposed normals, we adopt the idea from [6] which regards the mesh construction as an optimization problem. During each iteration, we utilize the generalized SfG [39] technique to firstly split current mesh into quadrangular faces according to the normal orientation of L_s and L_d, then stitch all the disconnected faces together to form a complete surface. The entire bas-relief modelling is proceeded in two steps: local shaping (split) and global blending (stitch) (shown in Fig. 9).

Let f_o denotes the expected output mesh, f_s and f_d denote the desired structure and detail, and h_f denotes the expected fixed overall height, then the total energy of bas-relief modelling is formed as:

$$E(f_o, f_s, f_d, h_u, h_f) = E_s(f_o, f_s) + \lambda_a E_d(f_o, f_d) + \lambda_b E_f(f_o, h_f) \qquad (6)$$

where $E_s(f_o, f_s) = \|f_o - f_s\|^2$ is the energy function that minimizes the difference between output faces f_o and structure faces f_s which aims to preserve the structure of transferred normal. $E_s(f_o, f_d) = \|f_o - f_d\|^2$ aims to preserve the details of transferred normal. $E_f(f_o, h_f) = \|h_{f_o} - h_f\|^2$ is to control the overall fixed height of relief. λ_a is the weight to recover the geometry details, and λ_b affects the style of the resulting bas-relief (roundness or flatness), and larger value of λ_b means flatter style.

6. Implementation details

In this paper, the normal computation and bas-relief modelling is implemented using C++ and OpenGL, and the normal transfer is implemented using Pytorch 1.3.1 with CUDA 10.0. All the experiments are performed on a desktop PC with two 2.10 GHz Intel(R) Xeon(R) Platinum 8160T CPU, 256 GB RAM and two NVIDIA TITAN
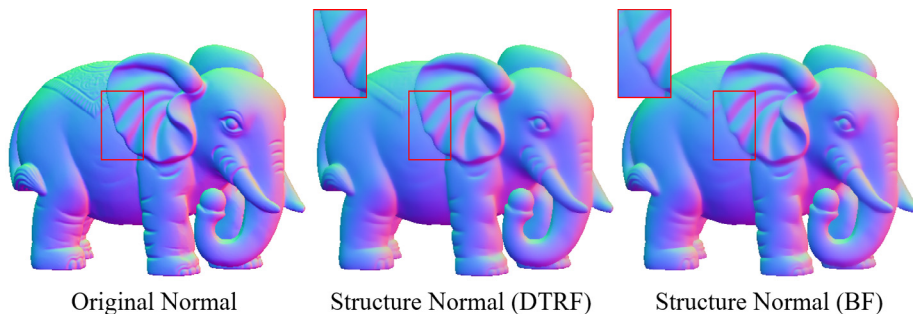

Fig. 7. Structure Normals obtained by DTRF $\{\sigma_s = 4, \sigma_r = 0.7\}$ and BF $\{\sigma_s = 3, \sigma_r = 0.3\}$. The zoom-ins of the red rectangle areas indicate that DTRF preserves better edges than BF in normal smoothing application.
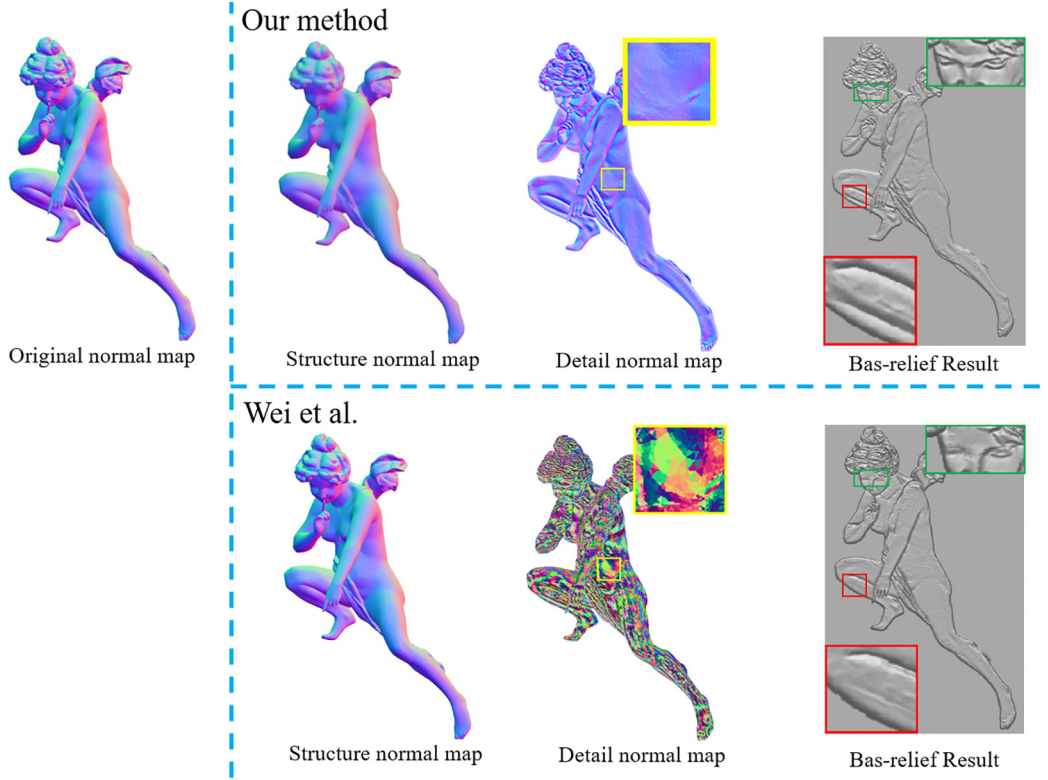
M. Wang, L. Wang, T. Jiang et al.

Fig. 8. Comparison of normal decomposition between Wei et al. [6] and ours. Our normal decomposition $\{\sigma_s = 10, \sigma_r = 1.0\}$ extracts details based on normal orientations and structure surface information as well, while Wei et al. [6] only obtain details which may damage the structure surface when detail enhancement is applied (see the zoom-ins of the red rectangle areas).
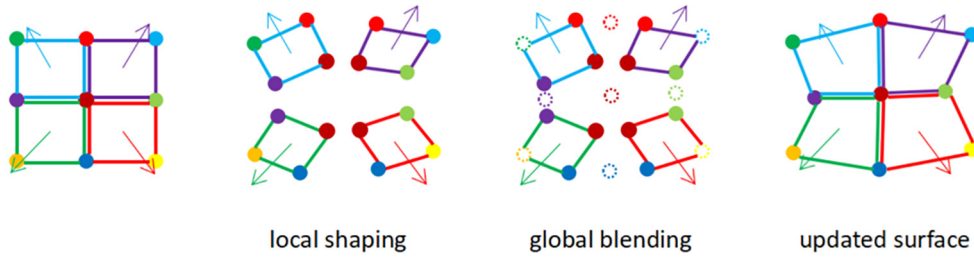

Fig. 9. The overflow of bas-relief modelling. In the local shaping step, each face (represented with the same colour edges and a normal vector) is projected according to its transferred normal vector. A vertex may be spit into two or four vertices, which are represented in the same colour. In the global blending step, new vertex positions (marked as hollow circles) are calculated by minimizing the total energy (Eq. (6)). These vertices are re-organized in originally connected way to form an updated surface. Then, we repeat the iteration on the updated surface until it is converged.

RTX graphics card. The Adobe Photoshop version used in this work is Photoshop CS6. A user-friendly GUI is created. We tested our method on a set of models with detail enhancement, and transferred texture details to various models in different scales and arbitrary shapes which demonstrate capability and effectiveness of diverse styles.

6.1. Parameters

For mask generation, we use the pretrained models of [10,12] provided by authors. Since the pretrained model of [12] is only trained on images with entire animal bodies, the segmented results for our source normals are not clear as target normals and could be arbitrary shapes like owl shown in Fig. 2. For normal transfer, our network has four parameters: $\{\alpha, \beta, \sigma, \kappa\}$ where they respectively control the structure preservation, texture preservation, texture scales and smoothness. Specifically, the ratio α/β presents the emphasis on either reconstructing the structures or the texture patterns. A larger ratio α/β indicates the structure identity of target normal in the synthesized result is strongly preserved, and a smaller ratio α/β indicates the texture patterns of source normal are effectively presented in the synthesized result. For a specific pair of source and target normals, user can adjust the trade-off between structure identity and texture patterns to create visually desired styles. For normal decomposition, our method has one user-specified parameter σ_s which is the weight of structure preservation.

For bas-relief modelling, our method has three parameters: $\{\lambda_a, \lambda_b, h_f\}$ where they are the weights of geometry preservation, flatness and fixed relief height, respectively. The normal image resolution for normal transfer and bas-relief modelling is fixed to 700×700. To reduce the number of user-specified parameters, we fix $\{\alpha = 1e0, \beta = 1e6, \sigma = 1e-3\}$ and place structure layers at {conv4_2} and texture layers at {conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1} which are default settings([9]) in normal transfer stage for all experiments.
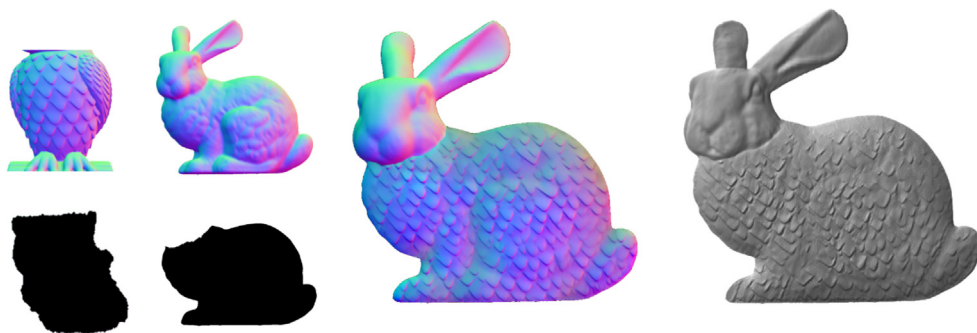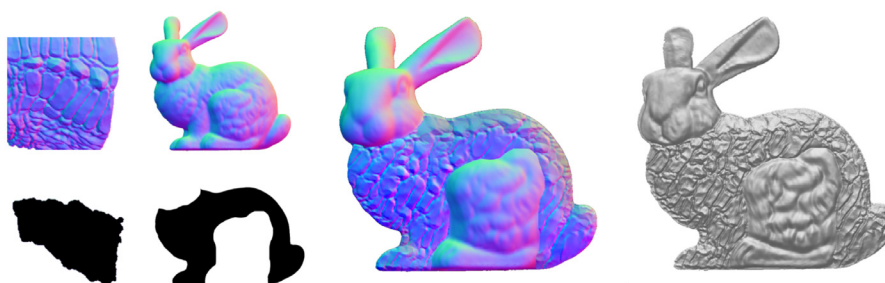
(a) Owl transfer on Bunny.
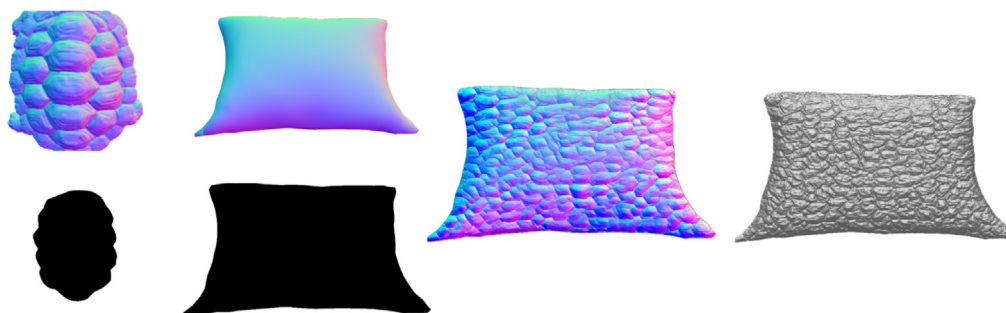
(b) Turtle transfer on Pillow.

Fig. 10. Examples of transfer results based on regular shapes of source and target normal images.

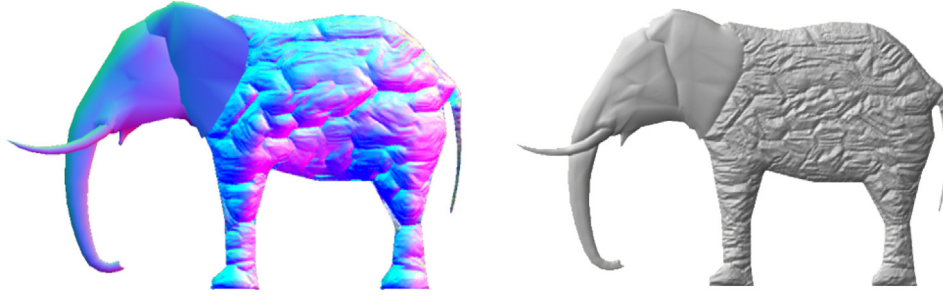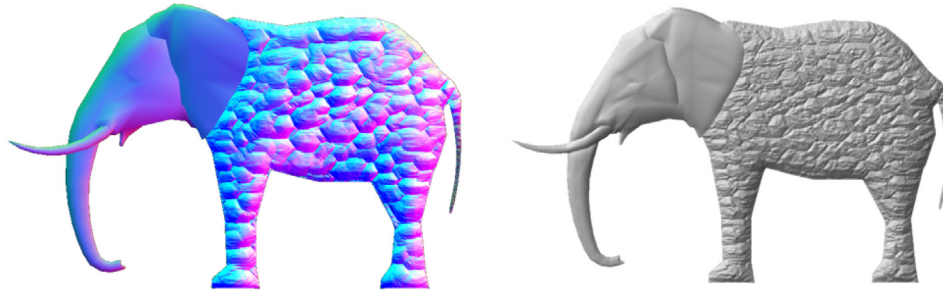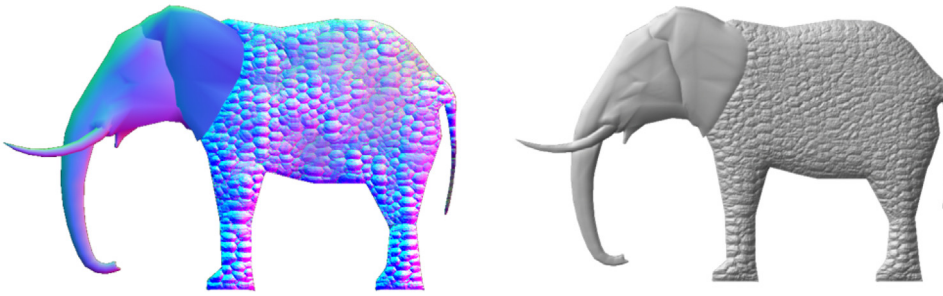(a) Turtle transfer on Buddha.

(b) Dragon transfer on Bunny.

Fig. 11. Example of transfer results based on arbitrary shapes of source and target normal images.

(a) Source normal and target normal.

(b) Turtle transfer on Elephant ($\kappa = 1.0$).

(c) Turtle transfer on Elephant ($\kappa = 0.5$).

(d) Turtle transfer on Elephant ($\kappa = 0.25$).

Fig. 12. An example of detail transfer results with multi-scale texture features.

7. Results and analysis

7.1. Detail transfer of bas-relief modelling

One advantage of our proposed method is the detail transfer on normal field, which is capable of transferring vivid texture patterns from one normal image to another. Unlike existing cut-and-paste

technology in [4], our approach learns the texture patterns from target normal images and transfers them onto source normal images in arbitrary shapes and multiple scales.

7.1.1. Regular shapes

Fig. 10 demonstrates that the proposed method transfers the partial texture details of Turtle shell onto the round belly shape
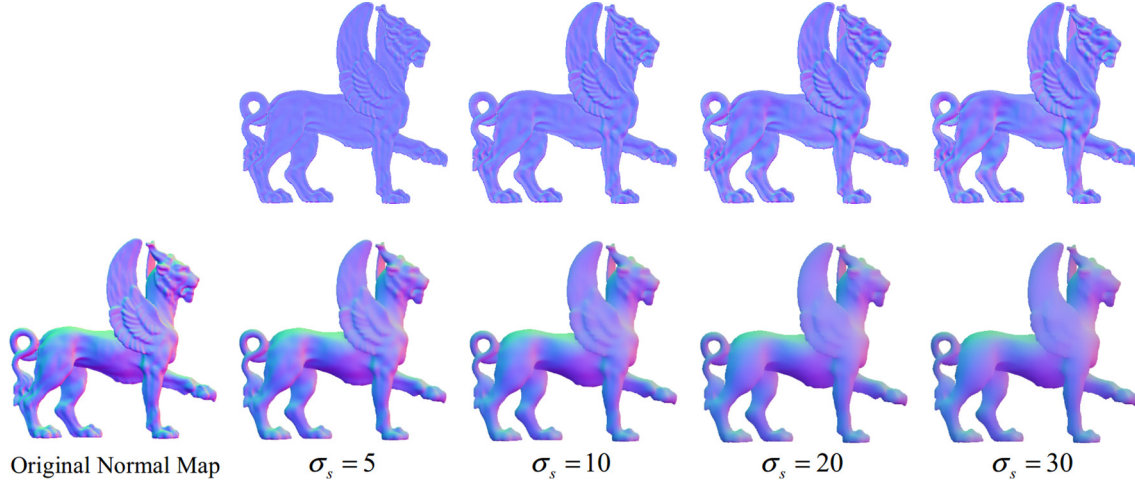
M. Wang, L. Wang, T. Jiang et al.

Fig. 13. Our normal decomposition results with different σ_s values. The first column is the original normal map as input. Results from the second column show the structure normals (bottom row) and corresponding detail normals (top row).
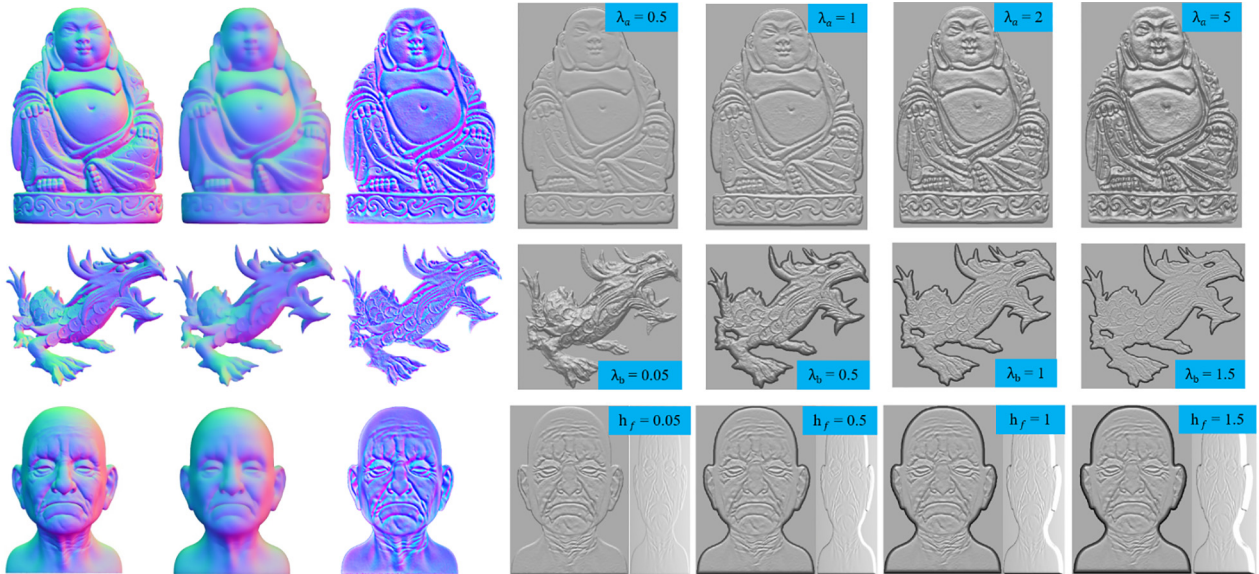

Fig. 14. Analysis of parameters λ_a, λ_b and h_f in Bas-relief modelling. In the left, the images are original normal map, structure normal map and detail normal map enhanced by Adobe Photoshop in each column. The first row shows that a larger λ_a value would enhance the detail preservation in relief result $\{\lambda_b = 0.5, h_f = 0.1\}$. The second row shows that a larger λ_b value would produce flatter relief results $\{\lambda_a = 1.5, h_f = 0.3\}$. The bottom row shows that a larger h_f value would increase the height of relief results $\{\lambda_a = 2, \lambda_b = 1.0\}$.

of Buddha and a rectangle-like shape of Pillow. The parameters are $\{\kappa = 0.4, \lambda_a = 2, \lambda_b = 0.05, h_f = 0.5\}$.

7.1.2. Arbitrary shapes

Besides the regular shape, our method still deals with arbitrary shapes as shown in Fig. 11. The texture details of the Owl and the Dragon are transferred to the body of the Bunny without any distortion. The parameters are $\{\lambda_a = 2, \lambda_b = 0.05, h_f = 0.5\}$. The masks of buddha and bunny are extracted in photoshop.

7.1.3. Multiscales

Our work is able to transfer different scales of details to the same target model by simply tuning one hyperparameter κ for texture scales during optimization process as shown in Fig. 12. The parameters $\{\lambda_a = 2, \lambda_b = 0.05, h_f = 0.5\}$.

7.2. Parameter σ_s for normal decomposition

In Fig. 13, we show the influence of parameter σ_s on our normal decomposition. As can be seen, the structure normal becomes smoother along with the increase of σ_s which preserves clearer details in detail normal. Thus, a larger σ_s value captures more details while preserving less structure information, and a lower σ_s value ignores some details while preserving more structure information. We choose the σ_s values between 10 and 20 in all of our experiments.

7.3. Hyperparameters for bas-relief modelling

In Fig. 14, we show how to tune the hyperparameters in our bas-relief modelling to get desired visual effects. For geometry preservation, a larger λ_a preserves surface details more clearly and even enhances it in an over-compressed case. A larger λ_b value
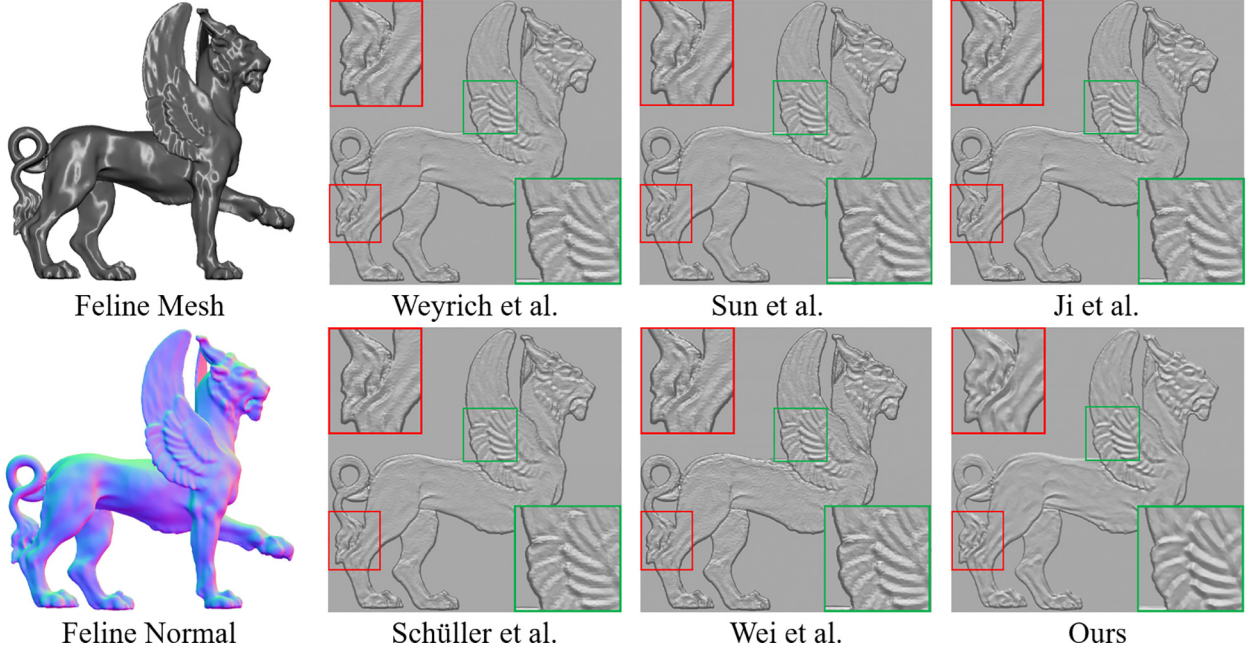
M. Wang, L. Wang, T. Jiang et al.

Fig. 15. Comparison to state-of-the-art methods with detail enhancement on standard thickness.The images in the first column is the original Feline mesh and Feline normal field, then the bas-relief results on the right columns follow the order: Weyrich et al. [40], Sun et al. [41], Ji et al. [4], Schüller et al. [15], Wei et al. [6] and ours $\{\lambda_a = 5, \lambda_b = 1.5, h_f = 0.1\}$. We recommend readers to view the electronic version for more clear details.
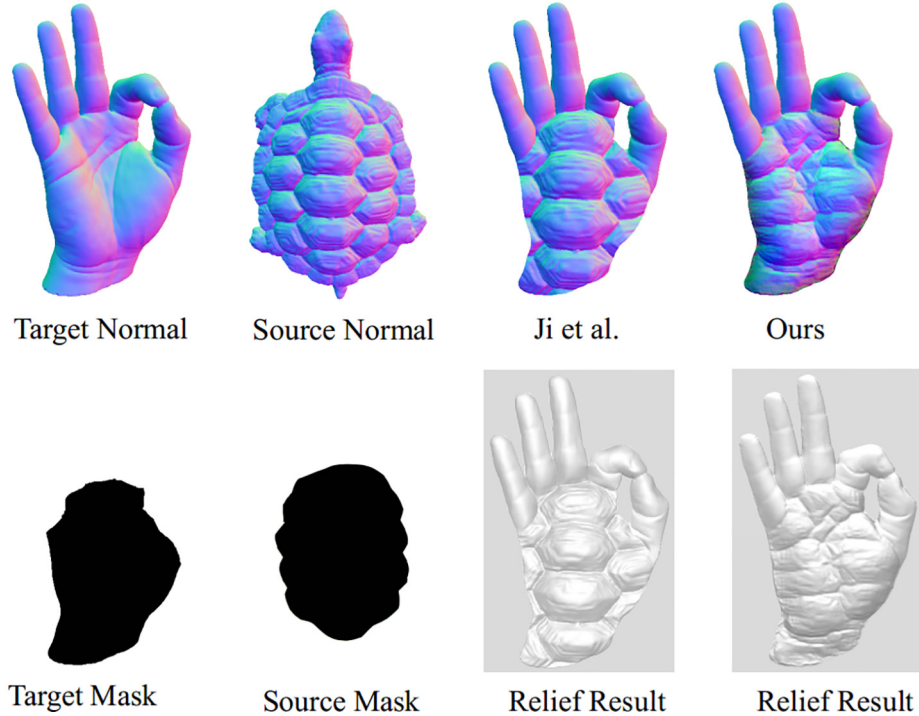

Fig. 16. Comparison on detail transfer between Ji et al. and ours $\{\lambda_a = 5, \lambda_b = 0.05, h_f = 3\}$.

produces a flatter bas-relief while a smaller one generates a round style of a model. Parameter h_f determines the overall height of the produced bas-reliefs.

7.4. Comparisons to previous literature methods

7.4.1. Comparison to state-of-the-art methods with detail enhancement

For fair comparison of the bas-relief results, a linear scaling as post-processing step is adopted from [6], which aims to make sure the generated bas-reliefs share same height since these approaches usually do not control their depth exactly. Additionally, we also carefully fine tune the hyper-parameters for each method and show their best visual appearance under the same lighting environment and height compression. We compare our approach with five state-of-the-art bas-relief modelling methods in a flatten style. All the produced results share some similarities, and details on Feline model are preserved well, which are shown in Fig. 15. However, our approach $\{\lambda_a = 5, \lambda_b = 1.5, h_f = 0.1\}$ produces more continuous and obvious edges (c.f. red

Table 1
Time consumption (seconds). Here shows the time occupation for typical bas-relief results.

Models	Stage 1 Normal transfer	Stage 2 Normal decomposition	Stage 3 Bas-relief modelling	Total
Buddha	110.78	16.36	47.98	175.12
Pillow	110.28	15.88	35.70	161.86
Bunny (Owl)	110.56	16.07	43.83	170.46
Bunny (Dragon)	109.08	16.15	43.86	109.09
Elephant (κ=1.0)	109.00	24.37	32.58	165.95
Elephant (κ=0.5)	109.57	19.58	36.25	165.40
Elephant (κ=0.25)	109.39	21.31	36.03	166.73

rectangles) and natural shape surface (c.f. green rectangles) than others.

7.4.2. Comparison on detail transfer between Ji et al. [4] and our method

Fig. 16 shows the detail transfer comparison between Ji et al. [4] and ours. Ji et al. [4] proposed a normal-based model method which constructed reliefs from normal images as well. However, their method utilized the cut-and-paste operation on the image domain to achieve detail transfer results which inevitably covered the original surface structures and details of target normal images. Our approach transfers the texture patterns to target normals while preserving original geometric property. As can be seen in Fig. 16, our result not only preserves well in the surface structures but also produces turtle textures spreading along with the detailed lines and surface of a human hand.

7.5. Time consumption

In general, our approach contains three stages which are normal transfer, normal decomposition and bas-relief modelling. The normal maps of 3D input meshes are fed into our semantic neural network in normal transfer stage. This stage generates synthesized normal images by running on two NVIDIA TITAN RTX graphics card. Next, we decompose the transferred normal result into structure normal and detail normal via our normal decomposition operator. Finally, the bas-relief modelling stage—surfaces are generated based on the normal images produced by previous stage. For typical models in this work, we record the time performance of three stages where stage 2 and 3 are implemented on CPU while stage 1 on GPU. Time cost of experiments in Fig. 10–12 can be found in Table 1.

7.5.1. Limitation

Our current semantic normal transfer method is a CNN network which regards the texture transformation as an online optimization problem (without any training process). It runs around 110 s in execution time for two 700×700 normal images which is slow in practice. To speed up the process, we recommend to replace the optimization process with feed-forward networks which may achieve real-time performance. The details on feed-forward networks can be found in these related works ([42–44]).

8. Conclusion

This paper presents a normal based bas-relief modelling method. To enrich the detailed features, we develop a semantic neural network of normal transfer which learns distributions of texture patterns and structure details from both source and target normal images respectively. Then a new normal image combining these distributions is generated by an optimization process. Unlike previous normal editing methods, our work is capable of learning the texture patterns from the source normal images and transfer-

ring them onto the target normal images in arbitrary shapes and multiple scales. To preserve geometric properties, we present a normal decomposition scheme which contributes to the generation of bas-relief results free from artefacts. A number of experimental results show that our method produces reasonable and pleasant bas-reliefs with enriched details and preserved geometry. Our future work will focus on speeding up the pipeline as our current semantic normal transfer network uses a slow optimization process. A promising solution is to use feed-forward networks instead of optimization process, which will save time for texture transformation.

CRediT authorship contribution statement

Meili Wang: Conceptualization, Methodology, Writing - original draft. **Li Wang:** Conceptualization, Methodology, Software, Writing - original draft. **Tao Jiang:** Software. **Nan Xiang:** Investigation, Software. **Juncong Lin:** Supervision, Writing - review & editing. **Mingqiang Wei:** Supervision, Writing - review & editing. **Xiaosong Yang:** Supervision, Validation, Writing - review & editing. **Taku Komura:** Supervision. **Jianjun Zhang:** Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

References

[1] T. Lawrence, Relief sculpture, School Arts Art Educ. Mag. Teach., 104, 2005, p. 3.
[2] Q. Zeng, R.R. Martin, L. Wang, J.A. Quinn, Y. Sun, C. Tu, Region-based bas-relief generation from a single image, Graph. Models 76 (3) (2014) 140–151.
[3] L. Governi, M. Carfagni, R. Furferi, L. Puggelli, Y. Volpe, Digital bas-relief design: a novel shape from shading-based method, Comput. Aided Design Appl. 11 (2) (2014) 153–164.
[4] Z. Ji, W. Ma, X. Sun, Bas-relief modeling from normal images with intuitive styles, IEEE Trans. Visual. Comput. Graph. 20 (5) (2013) 675–685.
[5] Z. Ji, X. Sun, W. Ma, Normal image manipulation for bas-relief generation with hybrid styles, arXiv preprint arXiv:1804.06092.
[6] M. Wei, Y. Tian, W.-M. Pang, C.C. Wang, M.-Y. Pang, J. Wang, J. Qin, P.-A. Heng, Bas-relief modeling from normal layers, IEEE Trans. Visual. Comput. Graph. 25 (4) (2018) 1651–1665.
[7] J. Shen, X. Yang, Y. Jia, X. Li, Intrinsic images using optimization, in: CVPR 2011, IEEE, 2011, pp. 3481–3487.
[8] J. Shen, X. Yang, X. Li, Y. Jia, Intrinsic image decomposition using optimization and user scribbles, IEEE Trans. Cybern. 43 (2) (2013) 425–436.

[9] L.A. Gatys, A.S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2414–2423.

[10] W. Wang, J. Shen, Deep visual attention prediction, IEEE Trans. Image Process. 27 (5) (2017) 2368–2378.

[11] W. Wang, J. Shen, H. Ling, A deep network solution for attention and aesthetics aware photo cropping, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2018) 1531–1544.

[12] X. Liang, X. Shen, D. Xiang, J. Feng, L. Lin, S. Yan, Semantic object parsing with local-global long short-term memory, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3185–3193.

[13] J. Kerber, M. Wang, J. Chang, J.J. Zhang, A. Belyaev, H.-P. Seidel, Computer assisted relief generation-a survey, in: Computer Graphics Forum, vol. 31, Wiley Online Library, 2012, pp. 2363–2377.

[14] Y.-W. Zhang, C. Zhang, W. Wang, Y. Chen, Z. Ji, L. Hui, Portrait relief modeling from a single image, IEEE Trans. Visual. Comput. Graph.

[15] C. Schüller, D. Panozzo, O. Sorkine-Hornung, Appearance-mimicking surfaces, ACM Trans. Graph. (TOG) 33 (6) (2014) 1–10.

[16] Z. Ji, X. Sun, S. Li, Y. Wang, Real-time bas-relief generation from depth-and-normal maps on gpu, in: Computer Graphics Forum, vol. 33, Wiley Online Library, 2014, pp. 75–83.

[17] Y.-W. Zhang, Y.-Q. Zhou, X.-L. Li, H. Liu, L.-L. Zhang, Bas-relief generation and shape editing through gradient-based mesh deformation, IEEE Trans. Visual. Compu. Graph. 21 (3) (2014) 328–338.

[18] Y.-W. Zhang, C. Zhang, W. Wang, Y. Chen, Adaptive bas-relief generation from 3d object under illumination, in: Computer Graphics Forum, vol. 35, Wiley Online Library, 2016, pp. 311–321.

[19] Y.-W. Zhang, Y. Chen, H. Liu, Z. Ji, C. Zhang, Modeling chinese calligraphy reliefs from one image, Comput. Graph. 70 (2018) 300–306.

[20] D. Zhang, X. Wang, J. Hu, H. Qin, Interactive modeling of complex geometric details based on empirical mode decomposition for multi-scale 3d shapes, Comput. Aided Design 87 (2017) 1–10.

[21] J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, F.C. Langbein, Y.-K. Lai, A.D. Marshall, Y.-H. Liu, Making bas-reliefs from photographs of human faces, Comput. Aided Design 45 (3) (2013) 671–682.

[22] J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, Y.-K. Lai, Y.-H. Liu, C. Wallraven, Use of non-photorealistic rendering and photometric stereo in making bas-reliefs from photographs, Graph. Models 76 (4) (2014) 202–213.

[23] H.T. To, B.-S. Sohn, Bas-relief generation from face photograph based on facial feature enhancement, Multimedia Tools Appl. 76 (8) (2017) 10407–10423.

[24] Z. Li, S. Wang, J. Yu, K.-L. Ma, Restoration of brick and stone relief from single rubbing images, IEEE Trans. Visual. Comput. Graph. 18 (2) (2011) 177–187.

[25] Y.-W. Zhang, J. Wu, Z. Ji, M. Wei, C. Zhang, Computer-assisted relief modelling: A comprehensive survey, in: Computer Graphics Forum, vol. 38, Wiley Online Library, 2019, pp. 521–534.

[26] D. Sýkora, L. Kavan, M. Čadík, O. Jamriška, A. Jacobson, B. Whited, M. Simmons, O. Sorkine-Hornung, Ink-and-ray: Bas-relief meshes for adding global illumination effects to hand-drawn characters, ACM Trans. Graph. (TOG) 33 (2) (2014) 1–15.

[27] T. Luo, J. Shen, X. Li, Accurate normal and reflectance recovery using energy optimization, IEEE Trans. Circ. Syst. Video Technol. 25 (2) (2014) 212–224.

[28] J. Shen, X. Yan, L. Chen, H. Sun, X. Li, Re-texturing by intrinsic video, Inf. Sci. 281 (2014) 726–735.

[29] T.Y. Wang, H. Su, Q. Huang, J. Huang, L.J. Guibas, N.J. Mitra, Unsupervised texture transfer from images to model collections, ACM Trans. Graph. 35 (6) (2016) 177–1.

[30] H. Huang, K. Xie, L. Ma, D. Lischinski, M. Gong, X. Tong, D. Cohen-Or, Appearance modeling via proxy-to-image alignment, ACM Trans. Graph. (TOG) 37 (1) (2018) 1–15.

[31] S. Berkiten, M. Halber, J. Solomon, C. Ma, H. Li, S. Rusinkiewicz, Learning detail transfer based on geometric features, in: Computer Graphics Forum, vol. 36, Wiley Online Library, 2017, pp. 361–373.

[32] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[33] W. Wang, J. Shen, L. Shao, Video salient object detection via fully convolutional networks, IEEE Trans. Image Process. 27 (1) (2017) 38–49.

[34] Z. Liang, J. Shen, Local semantic siamese networks for fast tracking, IEEE Trans. Image Process.

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[36] L. Gatys, A.S. Ecker, M. Bethge, Texture synthesis using convolutional neural networks, in: Advances in Neural Information Processing Systems, 2015, pp. 262–270.

[37] E.S. Gastal, M.M. Oliveira, Domain transform for edge-aware image and video processing, ACM SIGGRAPH 2011 Papers (2011) 1–12.

[38] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271), IEEE, 1998, pp. 839–846.

[39] W. Xie, Y. Zhang, C.C. Wang, R.C.-K. Chung, Surface-from-gradients: an approach based on discrete geometry processing, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2195–2202.

[40] T. Weyrich, J. Deng, C. Barnes, S. Rusinkiewicz, A. Finkelstein, Digital bas-relief from 3d scenes, ACM Trans. Graph. (TOG) 26 (3) (2007) 32–es.

[41] X. Sun, P.L. Rosin, R.R. Martin, F.C. Langbein, Bas-relief generation using adaptive histogram equalization, IEEE Trans. Visual. Comput. Graph. 15 (4) (2009) 642–653.

[42] D. Ulyanov, V. Lebedev, A. Vedaldi, V.S. Lempitsky, Texture networks: Feed-forward synthesis of textures and stylized images, in: ICML, vol. 1, 2016, p. 4.

[43] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, Springer, 2016, pp. 694–711.

[44] D. Chen, L. Yuan, J. Liao, N. Yu, G. Hua, Stylebank: an explicit representation for neural image style transfer, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1897–1906.

Meili Wang s an associate professor at College of Information Engineering, Northwest A&F University. She received her PhD degree in computer animation in 2011 at the National Centre for Computer Animation, Bournemouth University. Her research interests include computer graphics, geometric modeling, image processing, visualization and virtual reality.
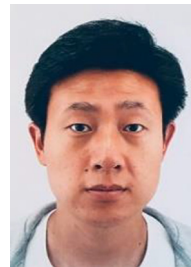
Li Wang received his B.Sci and M.Eng degrees from Department of Computer Science and Technology, Jilin University, China in 2013 and 2016, respectively. He is currently a PhD student at NCCA, Bournemouth University, UK. His interests include deep learning, computer vision and computer graphics.

Tao Jiang received the B.Sc. and M.Sc. degrees from the University of Electronic Science and Technology of China (UESTC). He is currently a Ph.D. student in the Faculty of Media & Communication at Bournemouth University. His research interests include computer vision and computer graphics, specifically, SLAM, shape deformation and registration.

Nan Xiang received his BS degree from the School of Software, Nanchang University, China, in 2012, the MA degree from School of Animation and Digital Arts, Communication University of China, in 2017. He is currently a PhD student at the National Centre for Computer Animation, Bournemouth University, UK. His research interests include virtual reality and 3D reconstruction.

Juncong Lin is currently a full professor in School of Informatics of Xiamen University, leading the Graphics and Virtual Reality Laboratory. He received his B.S. and Ph.D Degree both from Zhejiang University in 2003 and 2008 respectively. Before joining Xiamen University, he has worked in Chinese University of Hong Kong, JST ERATO IGARASHI Design Interface Project, Nanyang Technological University as a postdoc researcher. His research interests include shape modeling with creativity support, sketch based shape modeling and comic content processing. He has published over 20 papers including ACM SIGCHI, IEEE TVCG.

Taku Komura is a professor at the Institute of Perception, Action and Behavior, School of Informatics, University of Edinburgh. As the head of the Computer Animation and Robotics Group his research has focused on data-driven character animation, physically-based character animation, crowd simulation, cloth animation, anatomy-based modelling, and robotics. Recently, his main research interests have been in indexing and animating complex close interactions, which includes character-character interactions and character-object interactions.

Mingqiang Wei received his Ph.D degree (2014) in Computer Science and Engineering from The Chinese University of Hong Kong (CUHK). He is an associate professor at Nanjing University of Aeronautics and Astronautics (NUAA)'s School of Computer Science and Technology. Before joining NUAA, he served as an associate professor at Hefei University of Technology, and a postdoctoral fellow at CUHK. His research interest is computer graphics with an emphasis on smart geometry processing.

Jianjun Zhang is a professor of Computer Graphics at the National Centre for Computer Animation (NCCA), Bournemouth University where he leads the National Research Centre. Prof Zhang has published over 250 peer reviewed journal and conference publications and has given over 30 invited/keynote talks around the world. His research focuses on a number of topics relating to 3D Computer Animation, including virtual human modelling and simulation, geometric modelling, motion synthesis, soft body deformation and physics-based animation. He is also interested in virtual reality and medical visualization and simulation. He has chaired over 30 international conferences and symposia, and serves on a number of editorial boards.

Xiaosong Yang received his bachelor (1993) and master degree (1996) in Computer Science from Zhejiang University (P. R. China), Ph.D.(2000) in Computing Mechanics from Dalian University of Technology (PR China), PostDoc (2000–2002) in Tsinghua University on Scientific Visualization. He is currently an Associate Professor at the National Centre for Computer Animation, Bournemouth University. His research focuses on a number of topics relating to computer animation, texture synthesis, data mining, digital health, surgery simulation, etc.