Paper Review:Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression

Professor:
Doc. Manuel Loayza
by

Roxana Soto & Wilderd Mamani

Universidad Católica
San Pablo

Key Main

3D face reconstruction is the problem of recovering the 3D facial geometry from 2D images.This work is on 3D face reconstruction using only a single image.

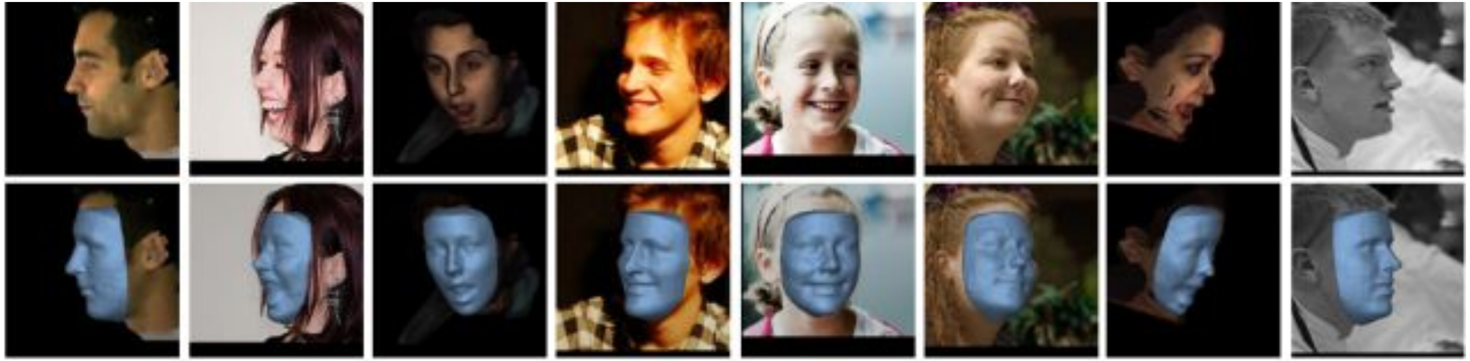


Figure 1: A few results from our *VRN - Guided* method, on a full range of pose, including large expressions.

Main Contributions

Based in 3D face reconstruction by using a novel volumetric representation of the 3D facial geometry, and an appropriate CNN architecture that is trained to regress directly from a 2D facial image to the corresponding 3D volume.

Methods

Dataset: http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm

Proposed volumetric representation:

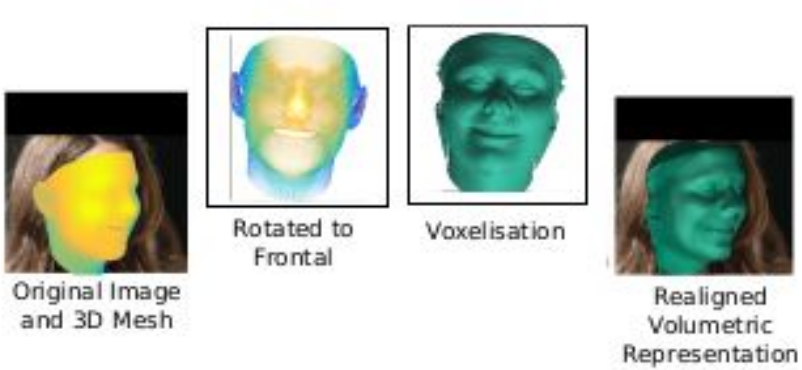

Original Image and 3D Mesh
Rotated to Frontal
Voxelisation
Realigned Volumetric Representation

Figure 2: The voxelisation process creates a volumetric representation of the 3D face mesh, aligned with the 2D image.

Methods

Volumetric Regression Networks

- **Volumetric Regression Network (VRN)**
  Our CNN architecture for 3D segmentation is based on the "**hourglass network"**

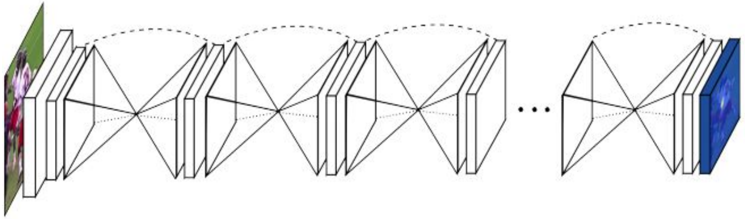


Fig. 1. Our network for pose estimation consists of multiple stacked hourglass modules which allow for repeated bottom-up, top-down inference.

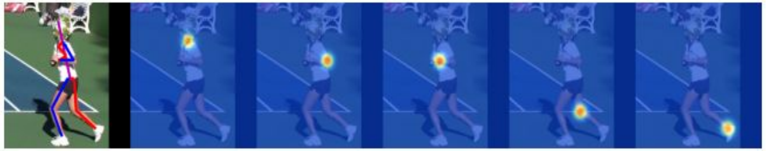


Fig. 2. Example output produced by our network. On the left we see the final pose estimate provided by the max activations across each heatmap. On the right we show sample heatmaps. (From left to right: neck, left elbow, left wrist, right knee, right ankle)

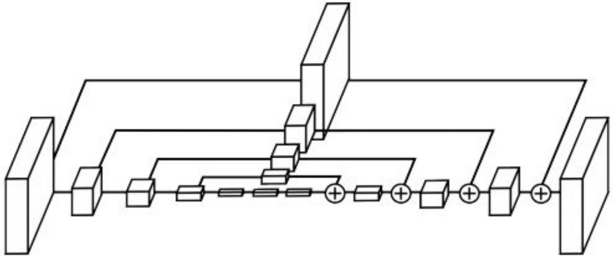
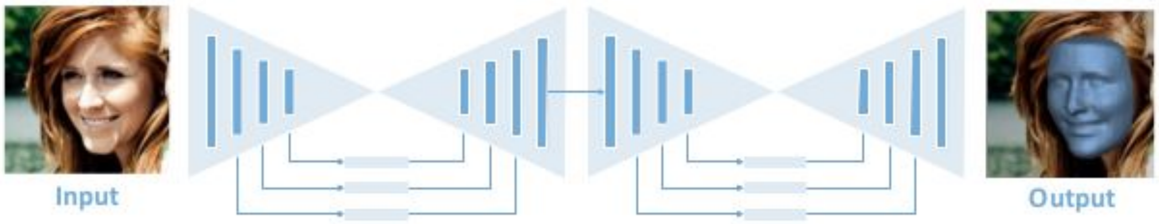


Fig. 3. An illustration of a single "hourglass" module. Each box in the figure corresponds to a residual module as seen in Figure 4. The number of features is consistent across the whole hourglass.

Methods

Volumetric Regression Networks

- **Volumetric Regression Network (VRN)**

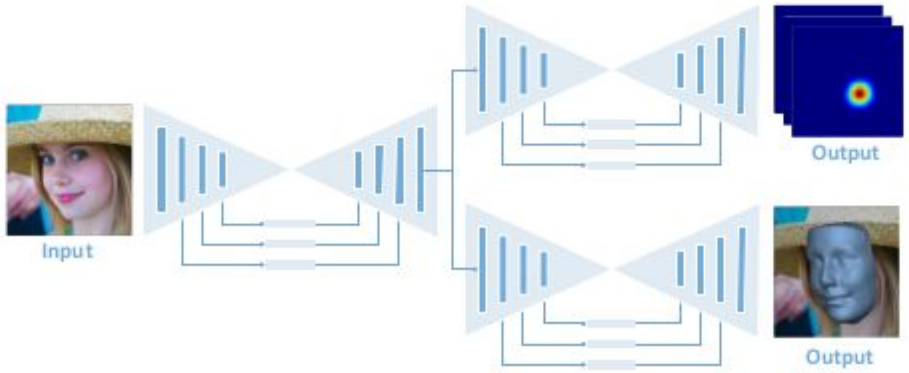


(a) The proposed *Volumetric Regression Network (VRN)* accepts as input an RGB input and directly regresses a 3D volume completely bypassing the fitting of a 3DMM. Each rectangle is a residual module of 256 features.

Methods

Volumetric Regression Networks

- **Volumetric Regression Network (VRN) - Multitask**

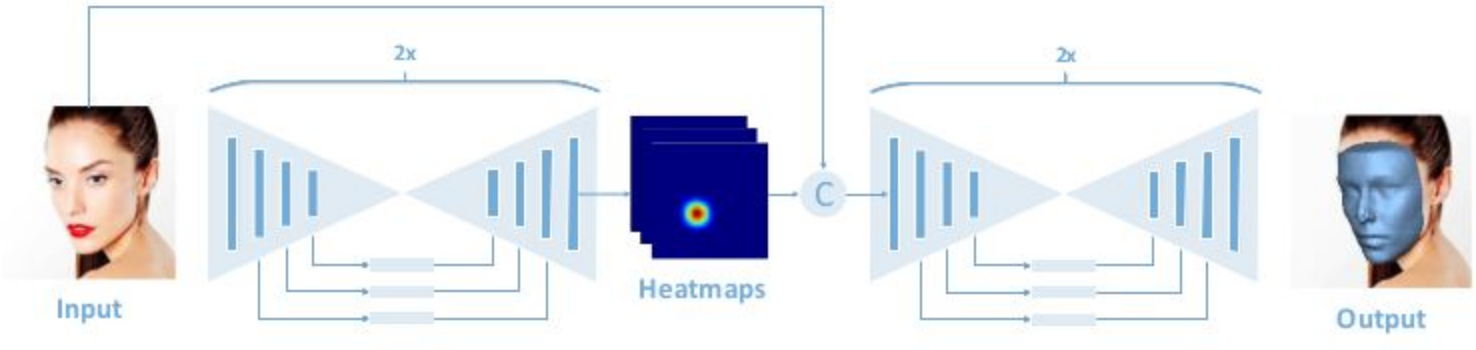


(c) The proposed *VRN* - *Multitask* architecture regresses both the 3D facial volume and a set of sparse facial landmarks.

Methods

Volumetric Regression Networks

- **Volumetric Regression Network (VRN) - Guided**



(b) The proposed *VRN - Guided* architecture firsts detects the 2D projection of the 3D landmarks, and stacks these with the original image. This stack is fed into the reconstruction network, which directly regresses the volume.

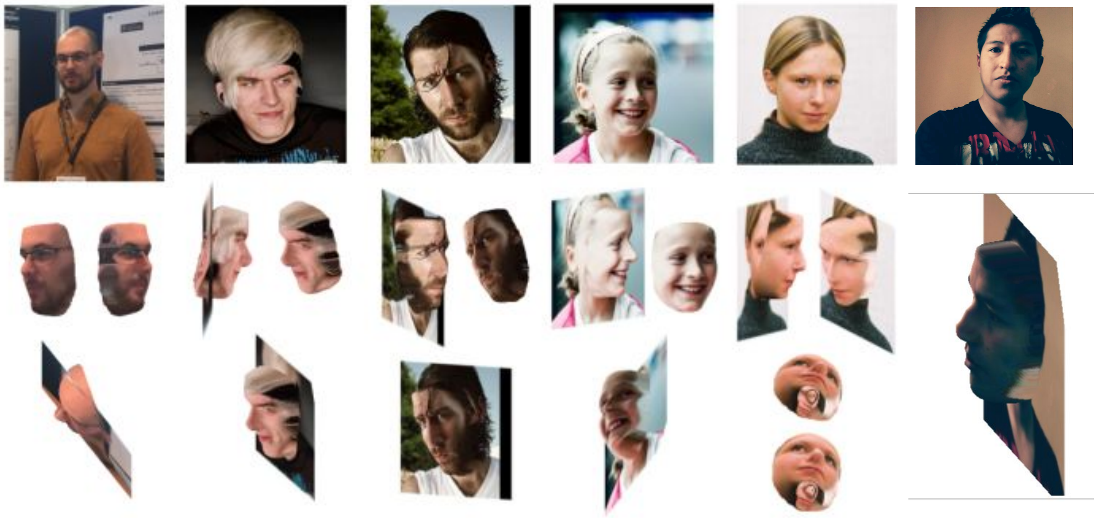
Training

Each of our architectures was trained end-to-end using **RMSProp** with an initial **learning rate of 10^-4,** which was lowered after 40 epochs to 10^−5

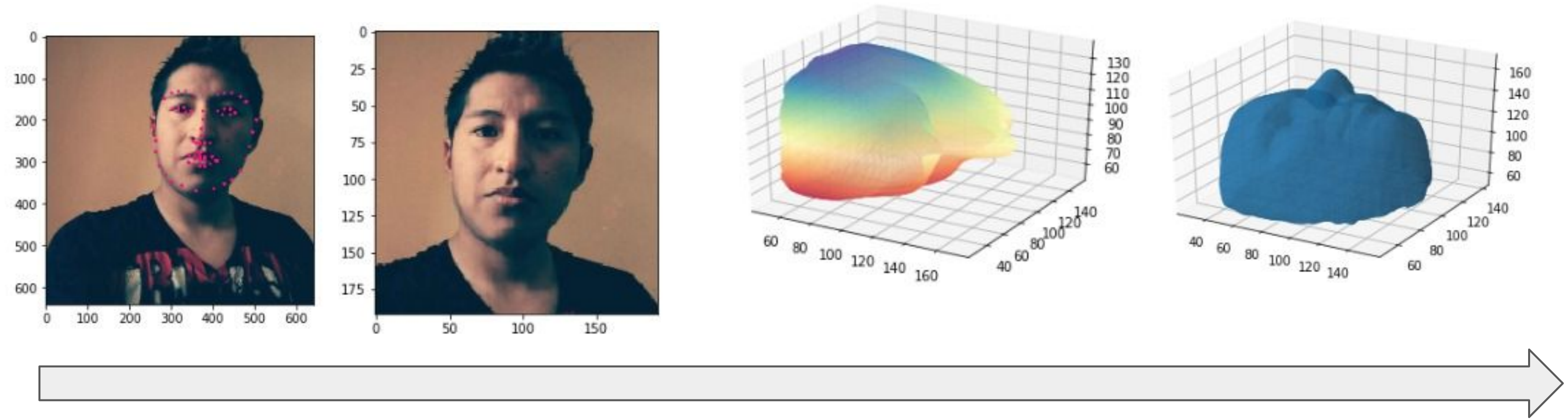
During training, **random augmentation** was applied to each input sample (face image) and its corresponding target (3D volume)with **rotation** and **translation** In 20 % of cases, the input and target were flipped horizontally. Finally, the input samples were adjusted with some colour scaling on each RGB channel. In the case of the **VRN - Guided**, the landmark detection module was trained to regress Gaussians with standard deviation of approximately 3 pixels (σ = 1).

Tests

http://cvl-demos.cs.nott.ac.uk/vrn/



Tests - processing the code in python



Results

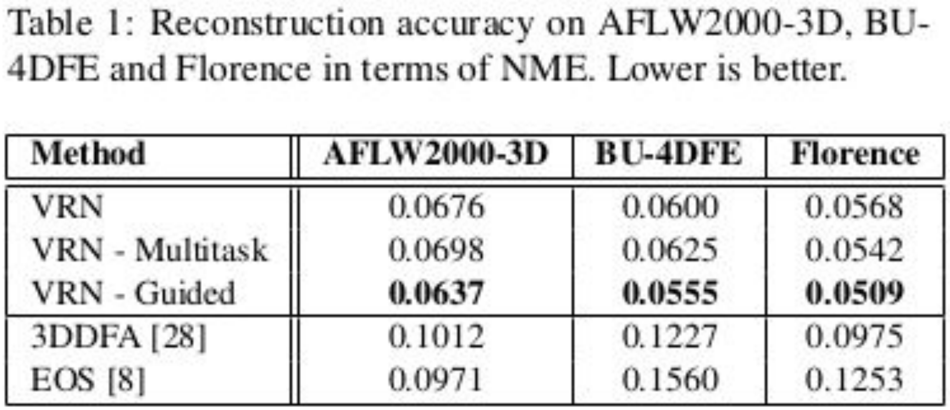


Table 1: Reconstruction accuracy on AFLW2000-3D, BU-4DFE and Florence in terms of NME. Lower is better.

Method	**AFLW2000-3D**	**BU-4DFE**	**Florence**
VRN	0.0676	0.0600	0.0568
VRN - Multitask	0.0698	0.0625	0.0542
VRN - Guided	**0.0637**	**0.0555**	**0.0509**
3DDFA [28]	0.1012	0.1227	0.0975
EOS [8]	0.0971	0.1560	0.1253