Statistical 3D Face Reconstruction with 3D Morphable Models

Claudio Ferrari, Stefano Berretti, Alberto Del Bimbo

claudio.ferrari@unifi.it

www.micc.unifi.it/3dmm-tutorial/

Department of Information Engineering (DINFO) &
Media Integration and Communication Center (MICC)

University of Florence (UNIFI), Florence, Italy

Deep Learning for 3D Face Modeling

Opening Remark

- In this part of the tutorial, we will present some recent deep learning based works that address the face reconstruction from single images problem and the basics of deep learning applied to 3D data
- We will focus mainly on those that employ the 3D Morphable Model in some way
- The goal is to give an overview of what has been done and what can be done
- A brief introduction to CNNs will be given, but a basic knowledge of the field (basic architectures, training procedures etc) is assumed

Deep Learning for 3D Vision

- The advent of deep learning techniques based on Convolutional Neural Networks (CNN) has drastically changed the way computer vision problems are being addressed;
- CNNs were specifically designed to work with 2D image data;
- The diffusion of such techniques in the 3D vision field has had a slower expansion because of (1) the **diverse data representation** i.e. irregular 3D data against regular RGB imagery and (2) **lack of 3D data**;
- Despite this, deep learning techniques are currently being applied with promising results also to 3D data

Representation Issue

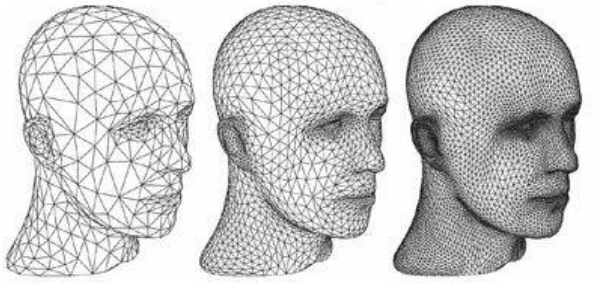
- 2D images: regular data structure

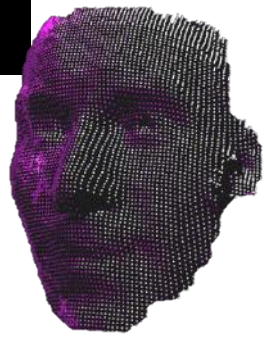
1	44	33	12	20	23	35	14
51	16	40	32	46	48	28	17
29	60	3	63	49	55	36	7
52	22	26	41	38	10	61	53
2	24	19	11	34	43	5	8
57	9	37	42	25	21	27	18
30	56	50	64	4	59	6	13
58	47	45	31	39	15	62	54

Can directly apply CNN

- 3D has many representations:
  - RGB(D) images or depth images
  - Point cloud
  - Polygonal Mesh
  - …

Need specific architectures and operators

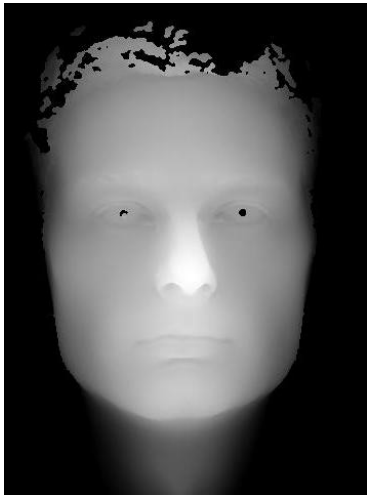


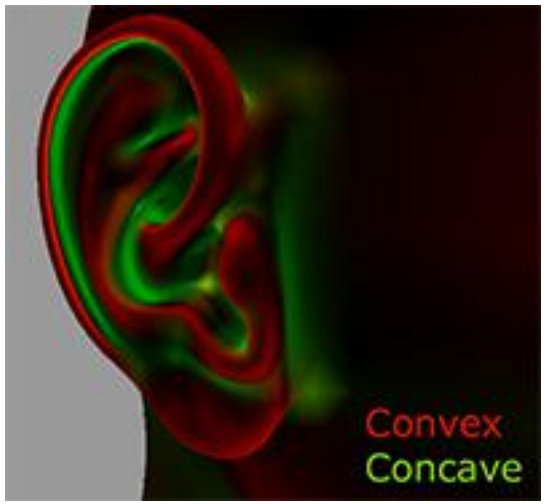


CNN for 3D data

- Geometric properties of surfaces can be encoded in 2D images:
  - Depth (z coordinate)
  - Normals
  - Curvatures
  - Composition (input tensors may have arbitrary dimensions…)

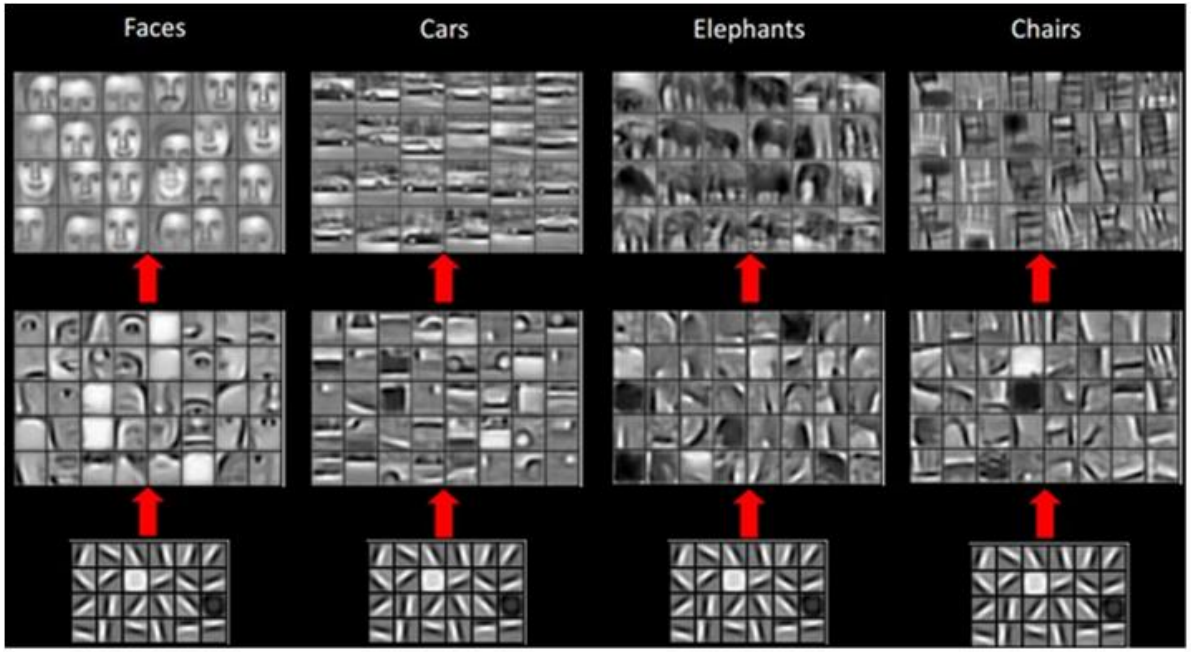
- This allows to directly apply a CNN to such data!





The problem of 3D data

- CNNs were rarely used in 3D vision until recently
- This was mainly due to the fact that CNN (more in general deep architectures) need A LOT of data to be effectively trained
- This because such networks learn lower and higher level abstractions directly from the raw image data → more data, better representation



The problem of 3D data

- Until few years ago, there was a lack of 3D data available
- Nowadays, millions of 3D models are available in online repositories (e.g. 3D Warehouse)
- Also, researchers put a lot of effort on the development of task-specific 3D dataset [1]
- However, a widely used workaround consists in exploiting existing techniques to produce synthetic data to be used for training (like the 3DMM …)

[1] Bulat, Adrian, et al.. "How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks)." *ICCV 2017*

A CNN for Regressing 3DMM Parameters

- We have seen that the first 3DMM performed a 3D face reconstruction by estimating a complex set of parameters to deform the 3DMM and render a synthetic image as similar as possible to the original

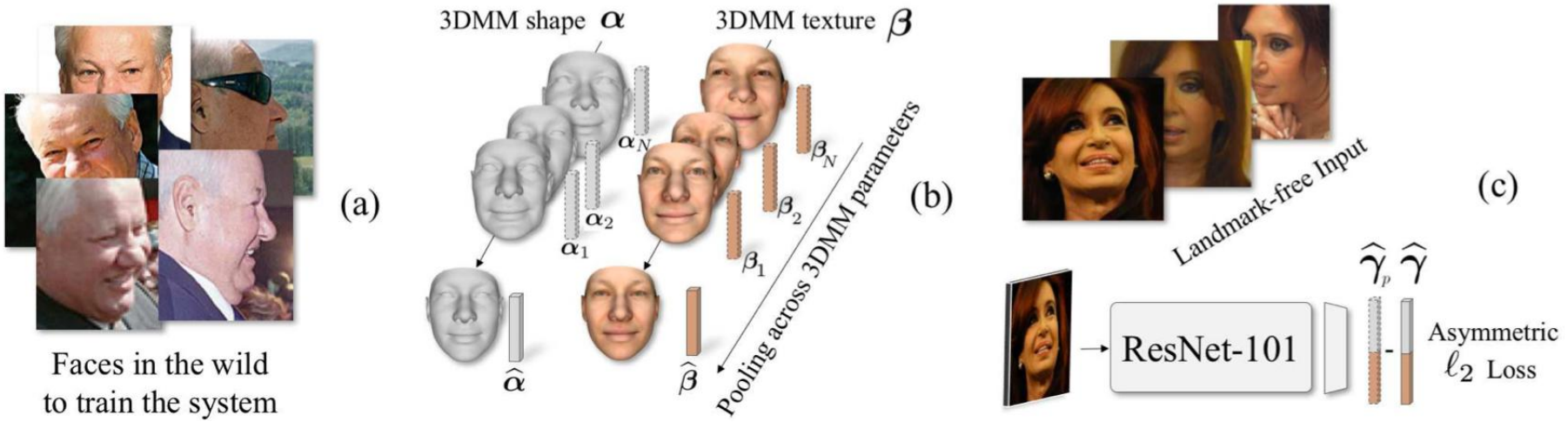
- So basically the problem was to find a mapping between pixels and the set of parameters…

- Inspired by this, Tran et al.[2] used a CNN to regress this mapping

[2] Tran, Anh Tuan, et al. "Regressing robust and discriminative 3D morphable models with a very deep neural network." *CVPR, 2017*

A CNN for Regressing 3DMM Parameters

- The 3DMM was applied to fit A LOT of 2D face images, clustered by identity
- For each identity, the 3DMM (shape and texture) parameters were pooled, assuming that the same subject should have the same set of parameters
- The parameters were used as signal to train the CNN and learn the mapping

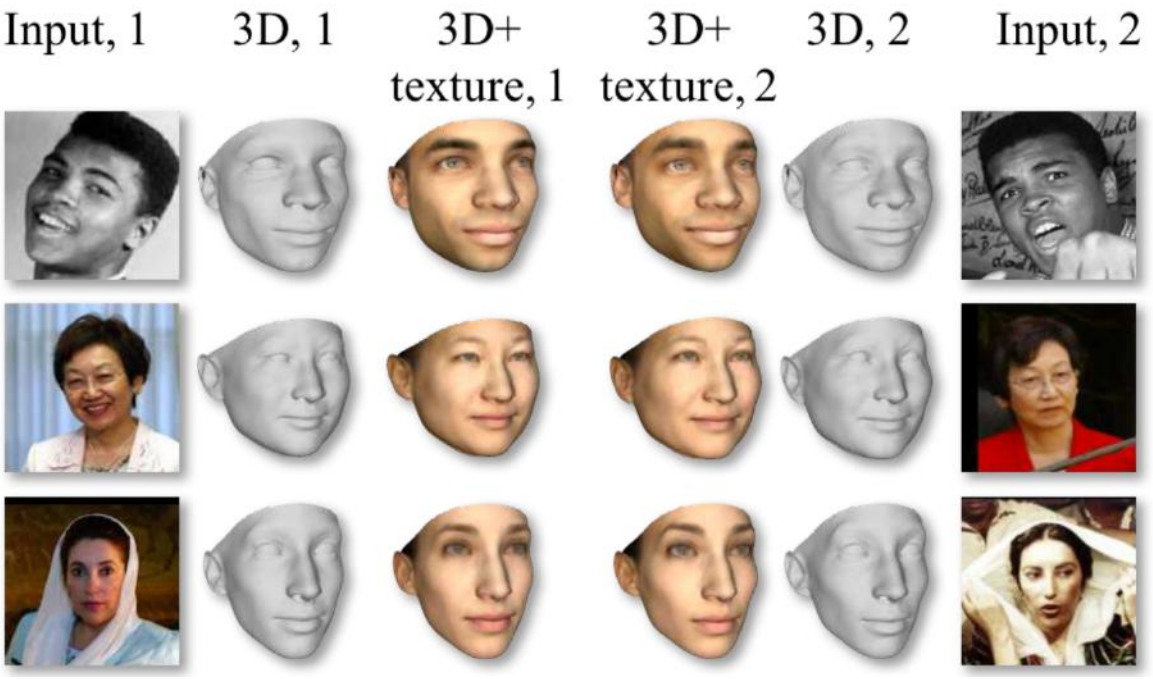


[2] Tran, Anh Tuan, et al. "Regressing robust and discriminative 3D morphable models with a very deep neural network." *CVPR, 2017*

A CNN for Regressing 3DMM Parameters

- The method demonstrated to be robust i.e. different images of the same subject led to very similar set of parameters

- The estimated parameters have been also used to perform “in the wild” face recognition, with promising results



[1] Tran, Anh Tuan, et al. "Regressing robust and discriminative 3D morphable models with a very deep neural network." *CVPR, 2017*

A CNN for Regressing 3DMM Parameters

- A limitation of the previous method is that expressions are not accounted for since the (1) the goal was to perform face recognition and (2) the 3DMM that was used i.e. the Basel Face Model (BFM) do not reproduce expressions

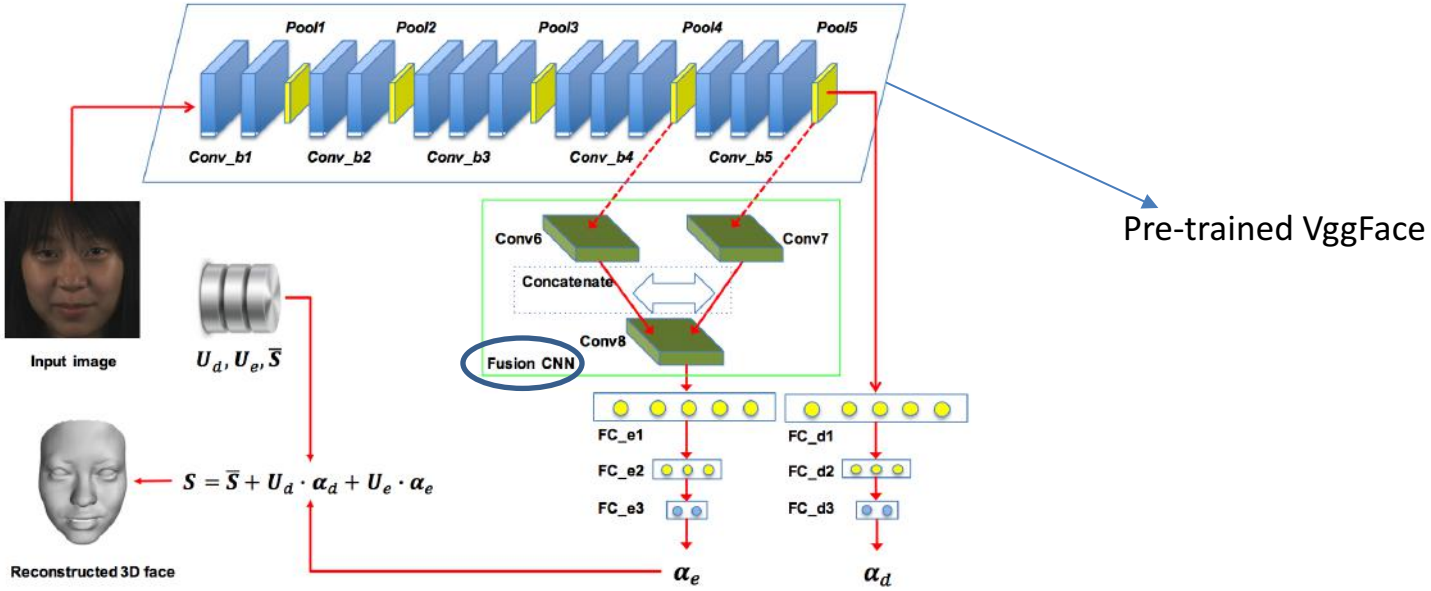
- A similar approach that is instead focused on 3D reconstruction and accounts for expressions is the one of Dou et al. [2]

- They use a combination of two models, the BFM to model global shape deformations and the AFM [3] model to account for expressions

[2] Dou, Pengfei, et al. "End-to-end 3D face reconstruction with deep neural networks." *CVPR*. 2017.
[3] Kakadiaris, Ioannis A., et al. "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach." *IEEE TPAMI 2007*

A CNN for Regressing 3DMM Parameters

- Similarly to the previous approach, the CNN takes an RGB image as input and try to regress the deformation parameters
- The loss function is a composition of the identity and expression parameters
- Higher level layers are class (identity) specific and thus suitable for identity parameters

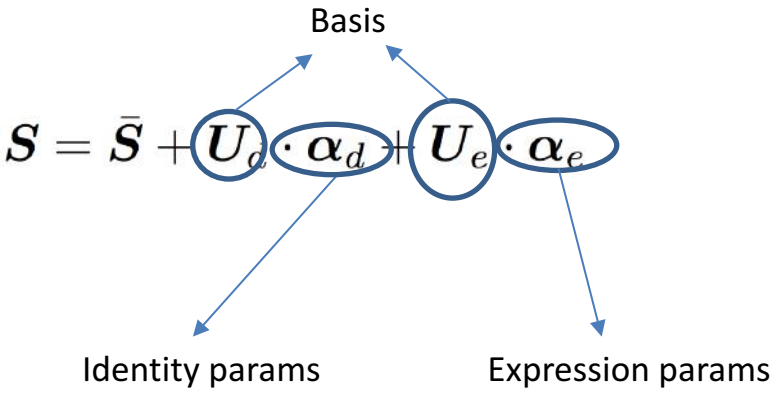


[2] Dou, Pengfei, et al. "End-to-end 3D face reconstruction with deep neural networks." *CVPR*. 2017.

A CNN for Regressing 3DMM Parameters

- Also in this case, the 3DMM is used to generate synthetic training data to be used in conjunction with real data to train the network

- The estimated parameters are used to generate the 3D shape associated to the image



End-to-end 3D Face Reconstruction

- The 3DMM framework has been (and still is) widely used to solve the 3D reconstruction problem

- CNN based solutions addressing the problem without including the 3DMM has been also developed (the training set, however, is built using a 3DMM…)

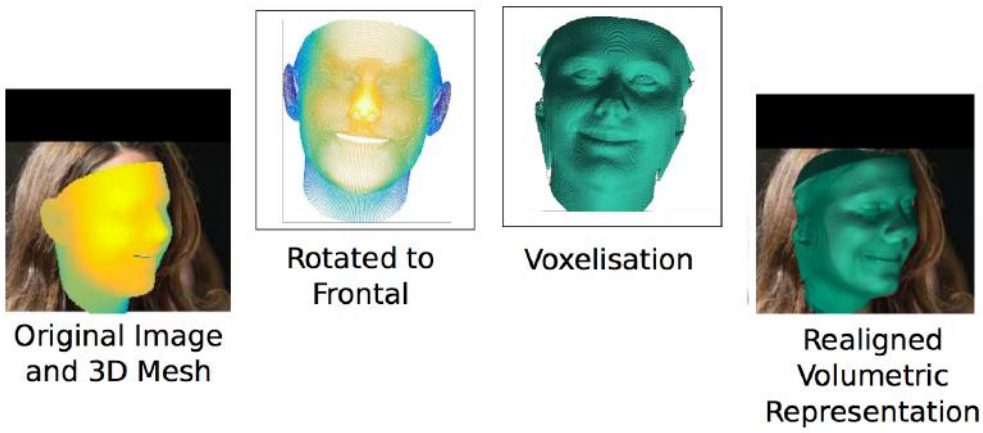
- An example is the work of Jackson et al.[4] in which a volumetric representation is regressed directly from the RGB image

- The output 3D model has at most the same resolution as the input image, however it bypasses many of the difficulties encountered in 3D reconstruction e.g. scans alignment, varying expressions etc

[4] Jackson, Aaron S., et al. "Large pose 3D face reconstruction from a single image via direct volumetric CNN regression." *ICCV* 2017.

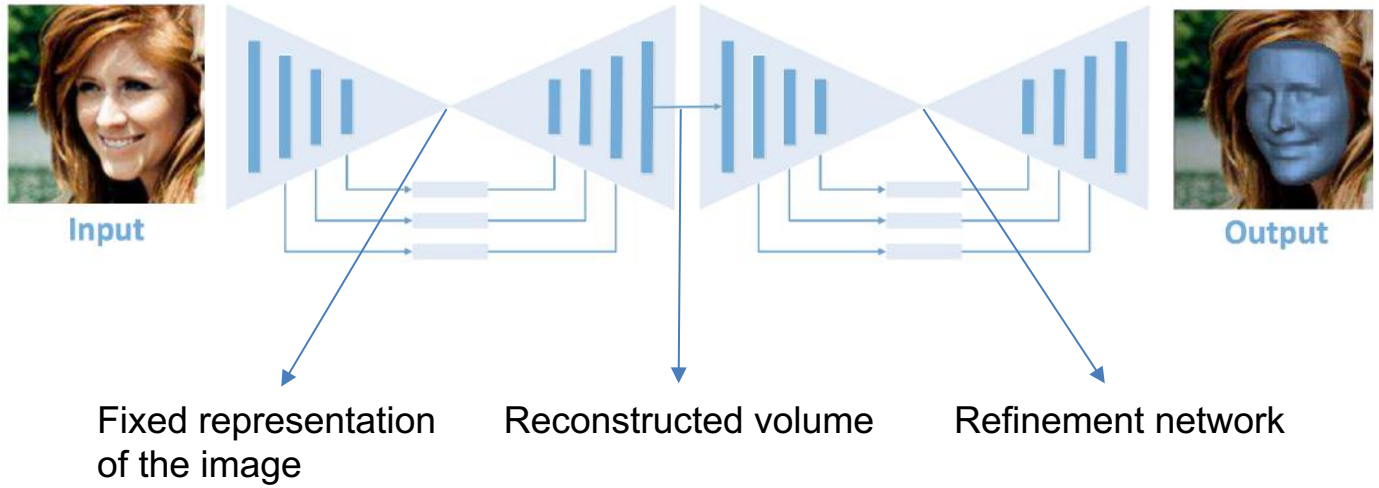
End-to-end 3D Face Reconstruction

- The main intuition is to derive a volumetric representation of the 3D models
- Scans are converted to a binary volume by discretizing the 3D space in voxels
- Points enclosed by the 3D scan are given value 1, otherwise 0
- 3D space discretized in a volume {h,w,d} of 192 X 192 X 200
- The goal is to learn a mapping from the image to volume $f : \mathbf{I} \rightarrow \mathbf{V}$



End-to-end 3D Face Reconstruction

- The architecture is composed of two encoder-decoder networks that allow to maintain the spatial consistency between input and output



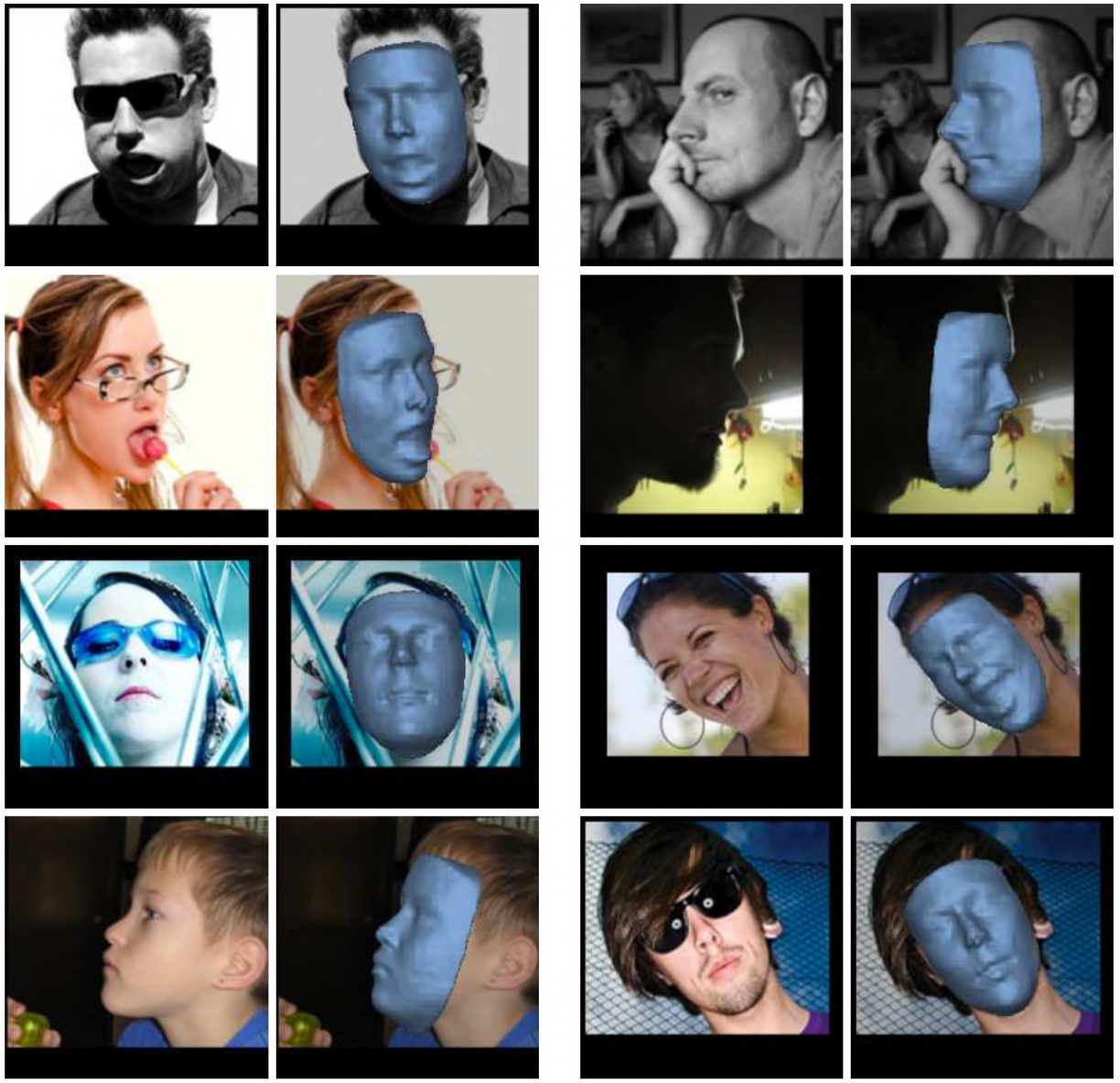
- The model is trained with a sigmoid cross entropy loss

Predicted value

GT value

$$l_1 = \sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{d=1}^{D}[V_{whd}\log\hat{V}_{whd} + (1-V_{whd})\log(1-\hat{V}_{whd})]$$

End-to-end 3D Face Reconstruction

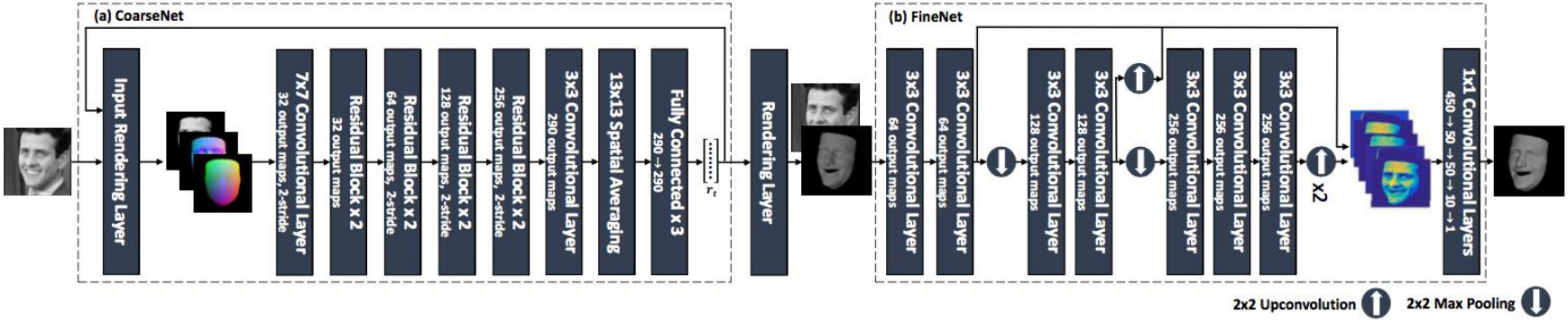


Fine Grained Details

- All the methods seen so far reconstruct a “smooth” foundation shape, usually without accounting for fine-grained details e.g. wrinkles
- While for some applications, a coarse geometry is sufficient, other may benefit from the recovery of fine details
- The 3DMM itself struggle in reproducing such fine details because of its intrinsic low dimensionality
- The power of CNNs can be exploited to recover highly detailed face surfaces

Fine Grained Details

- Richardson et al. [5] developed a network to reconstruct a detailed face shape from single image in a coarse-to-fine manner



- The first network (CoarseNet) is used to coarsely reconstruct the face shape

- FineNet is then used to generate fine details

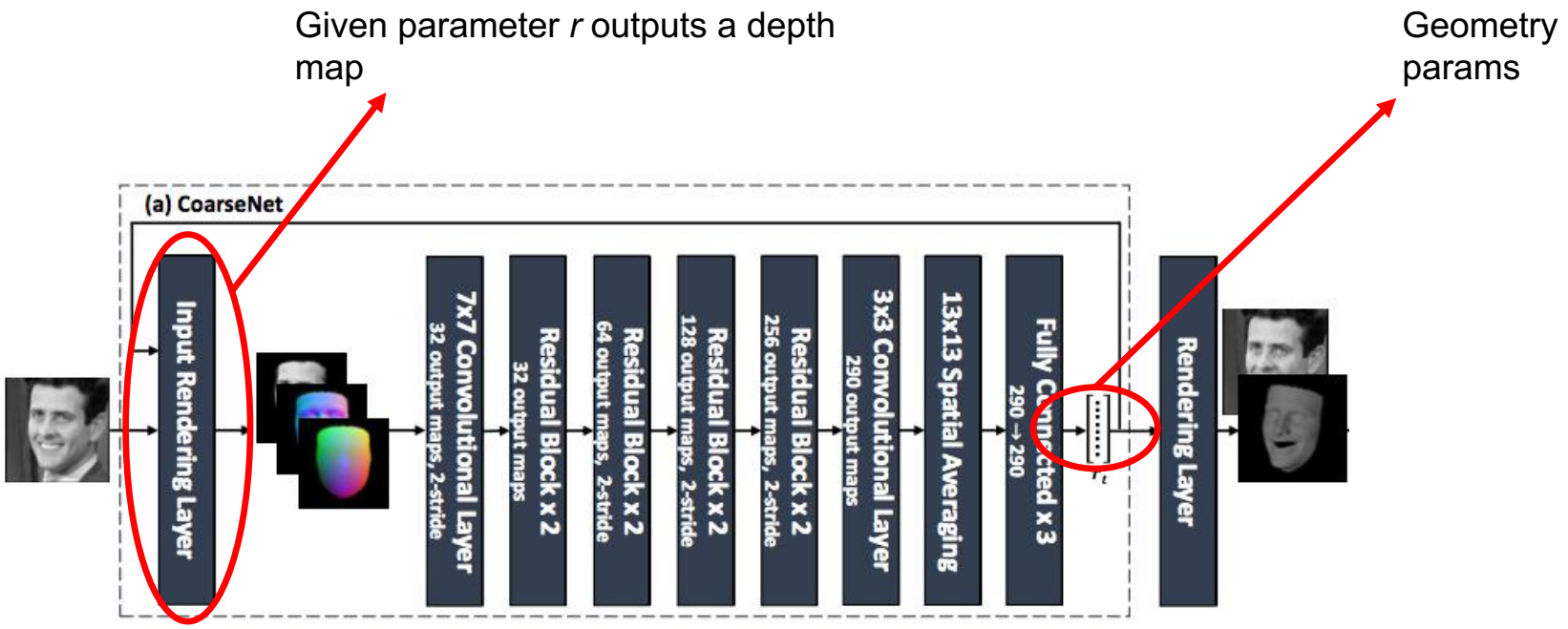
- Also in this case, the training set for the CoarseNet is synthetically generated using the 3DMM; the training set consists of a **set of images** with associated **3DMM parameters**

[5] Richardson E., et al. " Learning Detailed Face Reconstruction from a Single Image " *CVPR, 2017*

Fine Grained Details

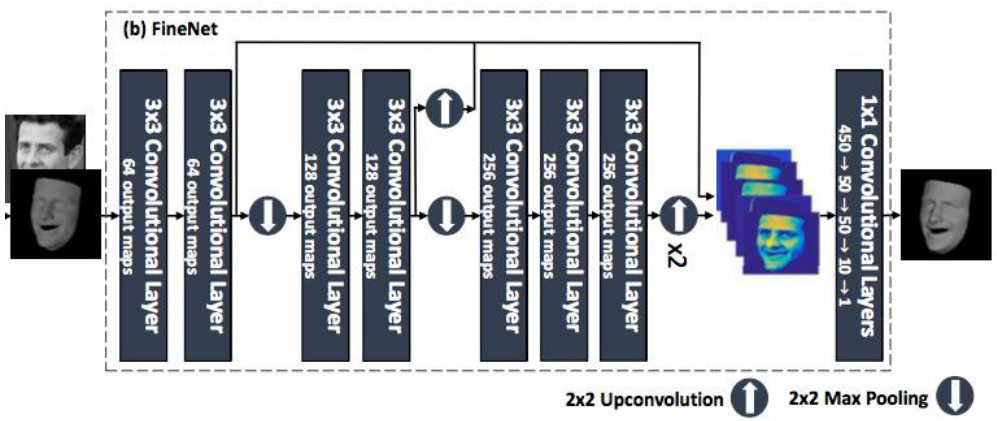
- Similarly to the previous works, CoarseNet is trained to regress the 3DMM and projection parameters (called *r*)

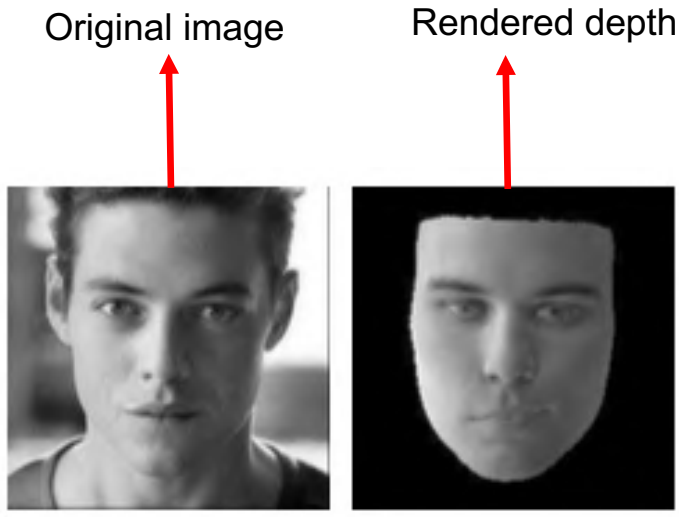
- These parameters are used by a rendering layer to render the input image with the current set of parameters *r*, (image-parameters pair is the actual input of the network)



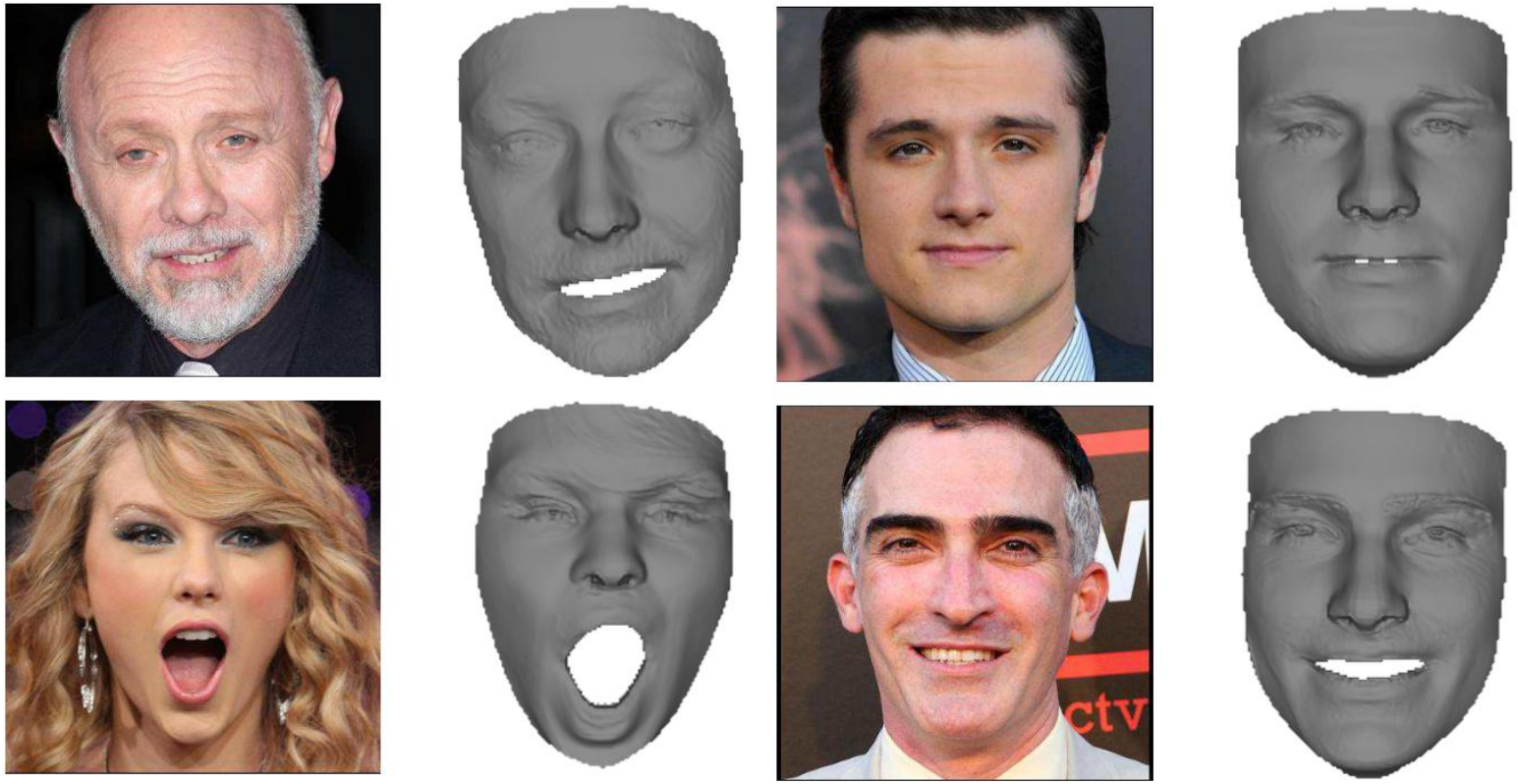
Fine Grained Details

- Finding (a sufficient number of) ground-truth detailed surfaces to train FineNet is not possible → **unsupervised training**
- The idea is to recover albedo and illumination parameters to render the recovered depth map
- FineNet is then trained to refine the depth map in order to match the appearance of the input image



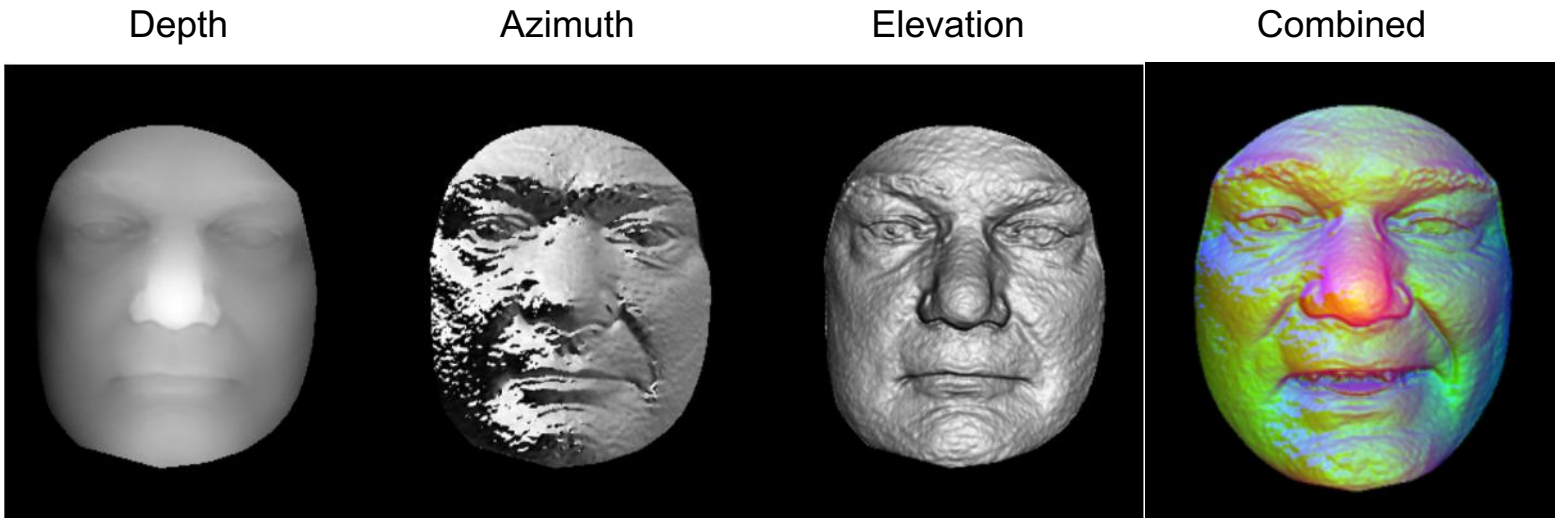


Fine Grained Details



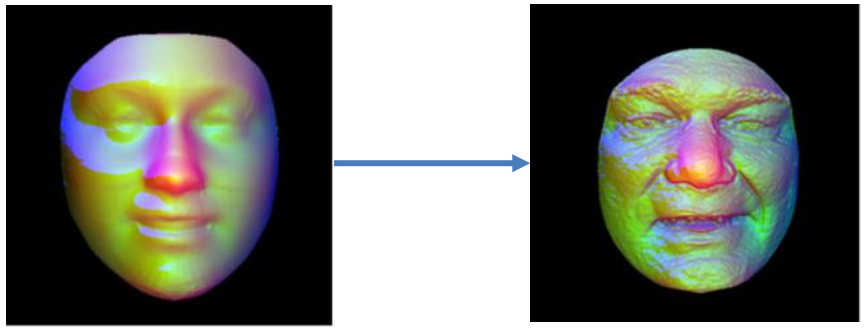
Fine Grained Details: Our Contribution

- Recently, we developed a system for refining the coarse reconstruction provided by a 3DMM
- It exploits a Conditional Generative Adversarial Network (CGAN)
- In our system, the 3D models are represented as RGB images where depth, azimuth and elevation angles of the surface normals constitute the 3 channels

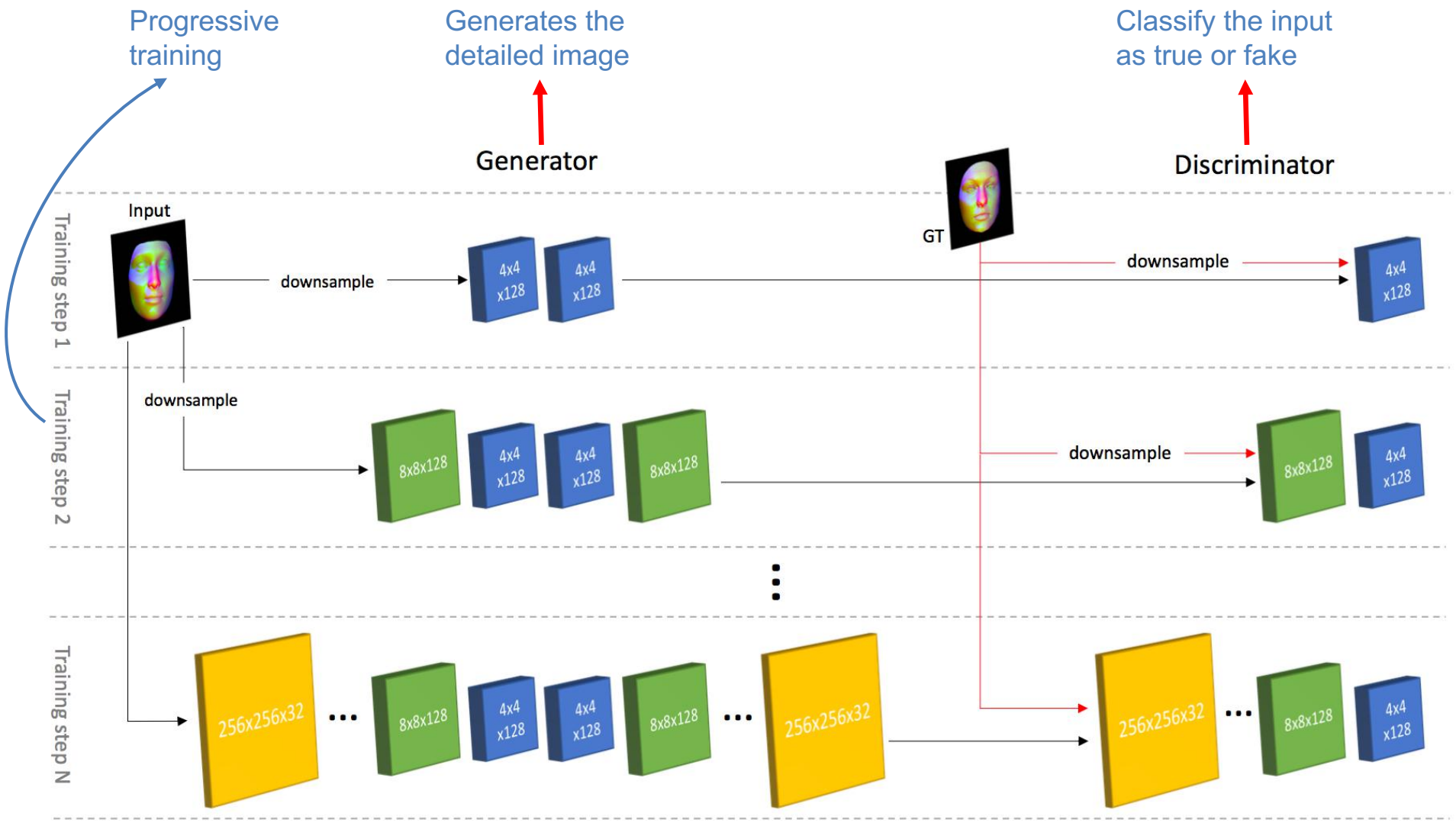


Fine Grained Details: Our Contribution

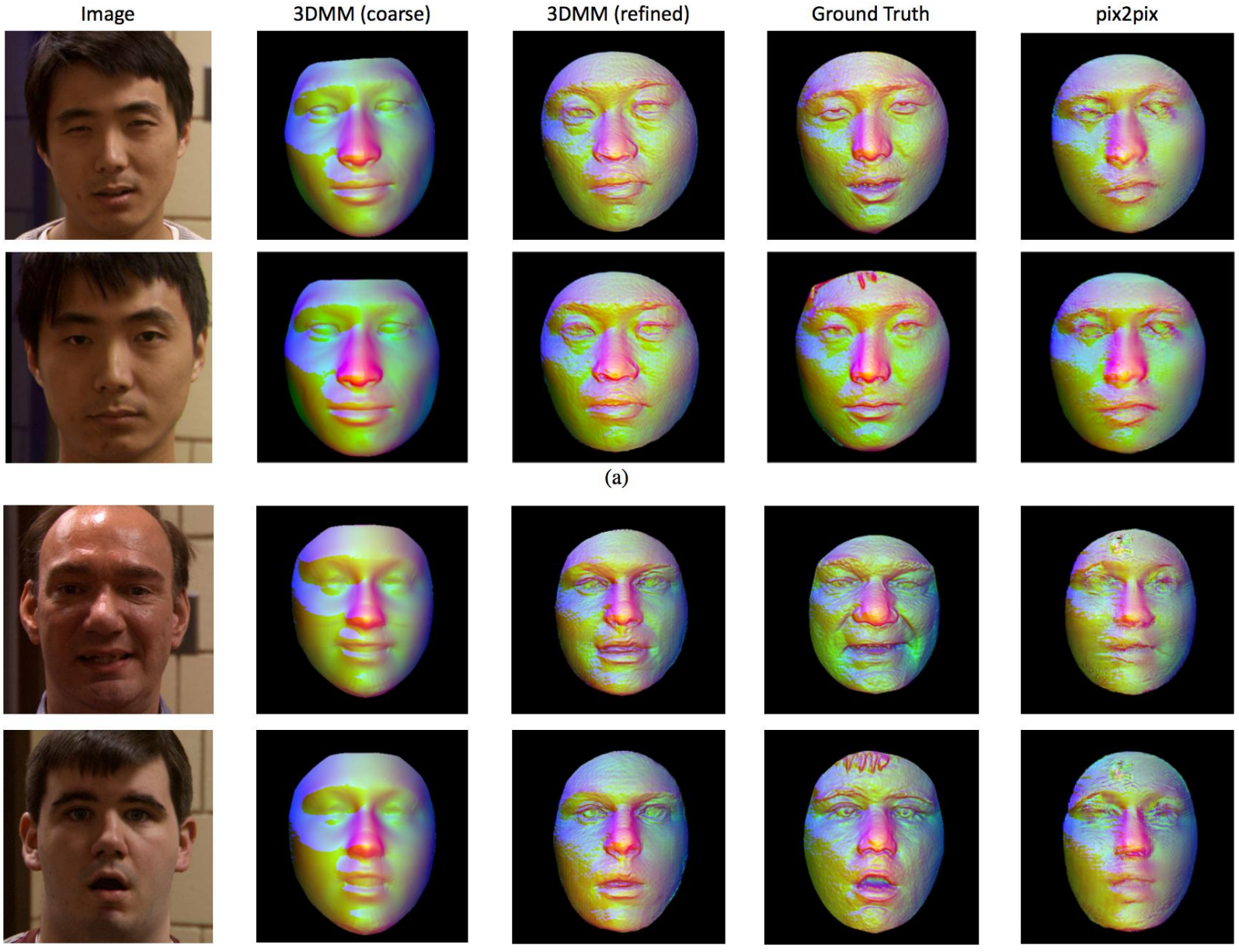
- The goal is to refine the coarse reconstruction of the 3DMM by adding fine grained details
- To train the system, the FRGC dataset was used (~4K scans of 400 subjects)
- The 3DMM is fit to the face images; the provided scans were used as ground truth
- No synthetic images were used; the system is trained on the “combined” images
- Augmentation based on random 3D rotations and scaling was applied
- Two different 3DMM were tested to assess the robustness of the approach i.e. our DL-3DMM and the Tran’s model



Fine Grained Details: Our Contribution



Fine Grained Details: Our Contribution



A Lot of Other Applications

- What we have seen so far is just a small part of what can be done with 3DMM and deep networks

- 3DMM generated images can be used with GANs to generate photorealistic face images (basically the problem is inverted)

Gecer B..., et al. " Semi-supervised Adversarial Learning to Generate Photorealistic Face Images of New Identities from 3D Morphable Model " ArXiv

- Large pose 2D (and 3D) landmark detection

Jackson A.S.., et al. " Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression " *ICCV, 2017*

- And many others

To Summarize

- We have revised some of the latest state-of-the-art applications that employ the 3DMM and deep networks
- There are many (many) others!
- Despite the great representational power of CNNs, the 3DMM is still an important tool that is mostly used to generate training data for the networks
- Clearly, to get the most out of CNNs, good reconstructions are needed; there is still room for new approaches!
- Huge amount of data needed!